Design Reference Guide

hp StorageWorks Continuous Access EVA V1.1B

Product Version: 1.1B

Seventh Edition (July 2004)

Part Number: AA-RS2YG-TE

This document is intended for system and network administrators who are experienced in SAN fabric configurations and in planning disaster-tolerant solutions. This document discusses design trade-offs and provides a checklist of required solution hardware and software components, based on the results of the design process.

For the latest version of this reference guide and other Continuous Access EVA documentation, access the HP storage website at http://h18006.www1.hp.com/storage/index.html.

hp invent

Printed in the U.S.A.

Continuous Access EVA V1.1B Design Reference Guide
Seventh Edition (July 2004)
Part Number: AA-RS2YG-TE

Contents

Figures

Tables

About This Guide

This guide is intended for all Continuous Access customers, existing and new, to provide a design or design review process of a Continuous Access EVA solution. The major purpose of these reviews is to answer the questions "will it work?" and "is it supported?" As such, this guide presents questions and outlines how to answer these questions. It also describes, at a high level, exactly what is and is not a Continuous Access EVA supported solution. It is intended more for managers and architects rather than operators, and does not contain procedures as much as an informal process for designing a Continuous Access solution that can be tailored to your company's needs.

This reference guide provides information to help you:

- Understand HP StorageWorks Continuous Access EVA features and benefits
- Become familiar with the design trade-offs and considerations
- Plan configurations

Topics include:

Revision history

Table 1: Revision history

Edition/Date	Summary
First (May 2003)	Initial publication

Table 1: Revision history (continued)

Edition/Date	Summary
Second (July 2003)	■ Increased configuration limits in the following: — Number of EVAs — Number of servers and FCAs per EVA ■ Added Microsoft® Windows® 2003 (32-bit) support ■ Added support for stretched Windows Clusters ■ Changed the Fibre Channel switch names ■ Added limited use configurations ■ Updated service pack support for Windows 2000
Third (August 2003)	■ Revised Vraid5 availability considerations ■ Added SLES 8 and Lifekeeper® support

Fourth (September 2003)	■ Added support for VCS V3.01 which includes support for:
	— EVA3000
	— 128 copy sets, 128 DR groups
	— New OS support:
	    — Windows 2003 64-bit
	    — HP HP-UX—11.23
	    — IBM AIX 5.2
	— Asynchronous replication
	— Multiple replication relationships
	— Removed gateway and WDM supported vendor lists
	— Increased separation distance (delay)

Fifth (December 2003)	■ Added OS support for: — Novell NetWare 6.5 — Linux (64-bit) ■ Increased number of FCAs to 256 per EVA
Sixth (December 2003)	■ Added Cisco switch support

Table 1: Revision history (continued)

Edition/Date	Summary
Seventh (July 2004)	■ Added support for: — OpenVMS 7.3-2 — Additional FCA support: — LP10000 — LP10000DC — VCS V3.02 — New boot from SAN and boot from CA replicated disk for HP-UX-11iv1 ■ Removed reference to HP Tru64 Stretched Clusters Configuration ■ Added chapter on Operating System Design Considerations ■ Renamed HP-UX-V11.11 to HP-UX-11iv1 ■ Renamed HP-UX-V11.23 to HP-UX-11iv2 ■ Added support for VCS V3.02 which changed the number of replication buffers from 32 to 64 ■ Added support for 250 GB FATA disk drives ■ Added support for 300 GB online disk

Intended audience

This document is intended for system and network administrators who are experienced in SAN fabric configurations and in planning disaster-tolerant solutions.

Related documentation

In addition to this document, HP provides the following related information:

- *HP StorageWorks Continuous Access EVA V1.1 Getting Started Guide*
- *HP StorageWorks Continuous Access User Interface V1.1A Installation Guide*
- *HP StorageWorks Continuous Access EVA V1.1B Operations Guide*
- *HP StorageWorks Continuous Access EVA V1.1B Release Notes*

- *HP StorageWorks Continuous Access User Interface V1.1A Release Notes*
- *HP StorageWorks Command View EVA V3.1 Getting Started Guide*
- *HP StorageWorks Enterprise Virtual Array User Guide EVA3000*
- *HP StorageWorks Enterprise Virtual Array User Guide EVA5000*
- *HP StorageWorks SAN Design Reference Guide*
- *HP StorageWorks Continuous Access and Data Replication Manager SAN Extensions Reference Guide*
- HP StorageWorks Continuous Access User Interface Online Help
- HP StorageWorks Command View EVA Online Help

Conventions

Conventions consist of the following:

- Document conventions
- Text symbols
- Getting help

Document conventions

The document conventions included in Table 2 apply in most cases.

Table 2: Document conventions

Element	Convention
Cross-reference links	Blue text: Figure 1
Key and field names, menu items, buttons, and dialog box titles	**Bold**
File names, application names, and text emphasis	*Italics*
User input, command and directory names, and system responses (output and messages)	`Monospace font` `COMMAND NAMES` are uppercase monospace font unless they are case sensitive
Variables	*`<monospace, italic font>`*
Website addresses	Blue, underlined sans serif font text: http://www.hp.com

Text symbols

The following symbols can be found in the text of this guide. They have the following meanings:

WARNING: Text set off in this manner indicates that failure to follow directions in the warning could result in bodily harm or death.

Caution: Text set off in this manner indicates that failure to follow directions could result in damage to equipment or data.

Note: Text set off in this manner presents commentary, sidelights, or interesting points of information.

Getting help

If you still have a question after reading this document, contact an HP authorized service provider or access our website: http://www.hp.com.

HP technical support

Telephone numbers for worldwide technical support are listed on the following HP website: http://www.hp.com/support/. From this website, select the country of origin.

Note: For continuous quality improvement, calls may be recorded or monitored.

Be sure to have the following information available before calling:

- Technical support registration number (if applicable)
- Product serial numbers
- Product model names and numbers
- Applicable error messages

- Operating system type and revision level
- Detailed, specific questions

HP storage website

The HP website has the latest information on this product, as well as the latest drivers. Access storage at http://www.hp.com/country/us/eng/prodserv/storage.html. From this website, select the appropriate product or solution.

HP authorized reseller

For the name of your nearest HP authorized reseller:

- In the United States, call 1-800-345-1518.
- In Canada, call 1-800-263-5868.
- Elsewhere, see the HP website for locations and telephone numbers: http://www.hp.com.

Continuous Access EVA Overview

1

This chapter describes the Continuous Access EVA solution, the basic features available in VCS and in particular Continuous Access EVA, and the benefits of using some of those features.

Topics include:

- Continuous Access EVA, page 15
- Features, page 19
- Other benefits, page 22

Continuous Access EVA

Continuous Access EVA is a Fibre Channel storage controller–based data replication (remote mirroring) solution to support disaster tolerance requirements. Continuous Access EVA works with the HP StorageWorks Enterprise Virtual Array (EVA) storage system, which contains the HSV virtualized RAID controller. The HSV controller and the Virtual Controller Software (VCS) version 3.00, 3.01, or 3.02 enhance the virtualization with remote replication technology.

Continuous Access EVA copies data online and in real time by using synchronous or asynchronous replication to a remote EVA through a local or extended storage area network (SAN). Additionally, data replication can be bidirectional, meaning that a storage array can be both a *source* and a *destination*. A particular LUN can be replicated in only one direction between the two storage arrays. Write I/O is sent to the source and then replicated by Continuous Access EVA to the destination. Properly configured, Continuous Access EVA can be a complete disaster-tolerant storage solution that guarantees data integrity if a storage system or site fails.

A basic Continuous Access EVA configuration is shown in Figure 1. Chapter 4 includes design recommendations for single switch and single fabric solutions. These recommendations are intended as entry level, proof of concept tests, or data distribution/migration solutions.

Additional information on data distribution and data migration can be found in the *HP StorageWorks Continuous Access EVA Operations Guide.* Chapter 4 of this guide also includes descriptions of advanced configurations such as multiple replication relationships and cascaded configurations.

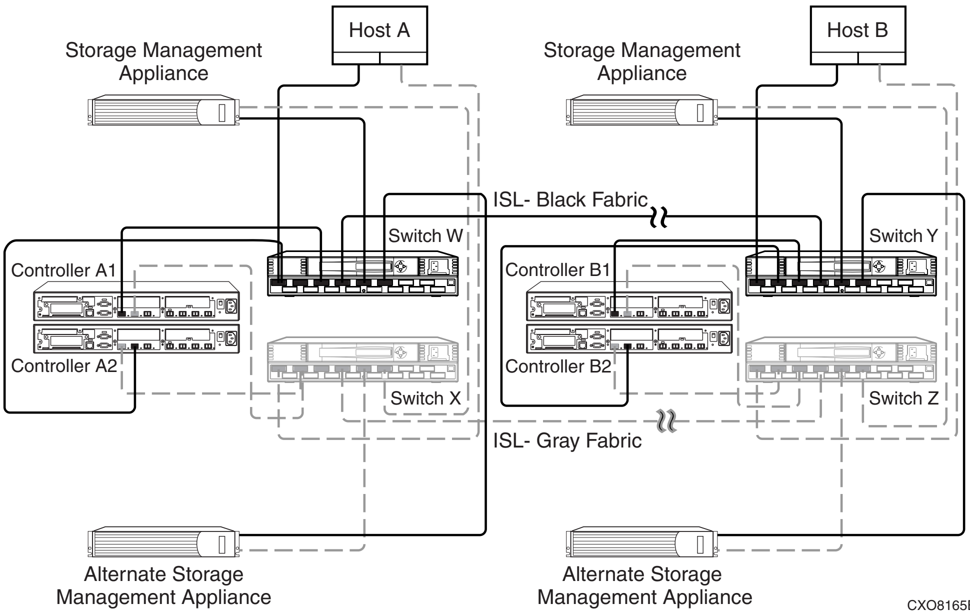

Figure 1: Continuous Access EVA basic configuration

HP OpenView Storage Management Appliance

The HP OpenView Storage Management Appliance (SMA) provides a centralized point for managing and monitoring SAN elements, including HP switches and storage arrays. It is located out of the data path and allows data transfers to proceed independently between servers and storage devices. The SMA eliminates the need for multiple terminal connections for managing and monitoring SAN elements. One SMA is required at each site; two at each site are recommended for no single point of failure (NSPOF) during a site failure. One SMA is needed to create a DR group, after which management services for both the source and destination storage arrays can be divided between two (or more) SMAs.

In some cases, the secondary appliance can be active and used to manage Continuous Access or non-Continuous Access functions on the secondary array.

Note: The active SMA must not be actively managing both EVAs in order to create a DR group or to add a member to a DR group.

For example, Business Copy for the EVA running on this secondary appliance may be needed to manage the creation of point-in-time copies of the destination copy of a copy set. As an active appliance, it is also ready to assume active management of one or both of the arrays if the primary appliance fails or to assist in site failover of the Continuous Access managed sets.

Supported operating systems and servers

Table 3 lists the operating systems and servers supported by Continuous Access EVA at the time of publication. For the most current list, contact your local HP Representative. For the supported versions of these operating systems and cluster support, refer to the *HP StorageWorks Continuous Access EVA Release Notes*.

Table 3: Continuous Access EVA–supported operating systems and servers

Operating System	Server
HP HP-UX—11.0	■ **A-Class:** rp24xx: A180, A180C, A400 (rp2400, rp2430), A500 (rp2450, rp2470), **B-Class:** B2600, **C-Class:** C3700, C3750, **J-Class:** J6750 **K-Class:** (64-bit only): Kx60, Kx70, Kx80, **L-Class:** rp54xx: L1000 (rp5400), L1500 (rp5430), L2000 (rp5450), L3000 (rp5470) **N-Class:** N4000 (rp7400), rp7410, rp8400 (HP-UX 11.11 only), **V-Class: (A5158A HBA only):** V2200, V2250, V2500, V2600 ■ **SuperDome (HP-UX 11.11 only) (16, 32, 64-way, PA-RISC processor versions only):** SD16000, SD32000, SD64000; PA8800 servers rp8420, rp7420, rp3440, rp4440, and SD-32/64/128
HP HP-UX—11i v1	■ **A-Class:** rp24xx: A180, A180C, A400 (rp2400, rp2430), A500 (rp2450, rp2470), **B-Class:** B2600, **C-Class:** C3700, C3750, **J-Class:** J6750 **K-Class:** (64-bit only): Kx60, Kx70, Kx80, **L-Class:** rp54xx: L1000 (rp5400), L1500 (rp5430), L2000 (rp5450), L3000 (rp5470) **N-Class:** N4000 (rp7400), rp7410, rp8400 (HP-UX 11.11 only), **V-Class: (A5158A HBA only):** V2200, V2250, V2500, V2600 ■ **SuperDome (HP-UX 11.11 only) (16, 32, 64-way, PA-RISC processor versions only):** SD16000, SD32000, SD64000; PA8800 servers rp8420, rp7420, rp3440, rp4440, and SD-32/64/128

Table 3: Continuous Access EVA–supported operating systems and servers (continued)

Operating System	Server
HP HP-UX—11i v2	■ rx1600, rx5670, rx2600, zx2000, zx6000, rx4640, rx7620, rx8620, Superdome Integrity ■ No support for MX2 processor based servers until late May 2004
HP OpenVMS—V7.2-2, V7.3-1, V7.3-2 HP Tru64 UNIX— V5.1a, and V5.1b	■ AS800, AS1200, AS4000, AS4100, AS8200, AS8400, DS10, DS10L, DS15, DS20, DS20E, DS20L, DS25, ES40, ES45, ES47 (PCI-X), ES80 ■ (PCI-X), GS60, GS60E, GS80, GS140, GS160, GS320, GS1280 (PCI-X)
IBM AIX—V4.3.3, V5.1, and V5.2	■ P610, P615, P620, P630, P640, P650, P655, P660, P670, P680, P690, F50, F80, H50, H70, H80, M80, S7A, S70, S80
Microsoft® Windows® Server 2003 (32- and 64-bit) Microsoft Windows 2000 Advanced Server Microsoft Windows NT®	■ Intel and ProLiant x86 Servers, ProLiant BL20P, BL40P (W2K, WS2003 only)
Novell NetWare—V5.1, V6.0, and V6.5	■ Intel and ProLiant x86 Servers
Red Hat Linux—Advanced Server V2.1 (32-bit)	■ ProLiant x86 Servers, ProLiant BL20P, BL40P (B-series and M-series switches)
Red Hat Linux—Advanced Server V2.1, 3.0 (64-bit)	■ rx5670, zx6000, rx4640, rx1600
Sun Solaris—V2.6, V7, V8, V9	■ Sun Fire 280R, v210, v240, v440, v480, v880, v1280 (PCI only); Sun Fire 3800 (cPCI only) ■ Sun Fire 3800 (cPCI only) ■ Sun Fire 4800, 4810, 6800 (PCI and/or cPCI, with I/O Assembly change) ■ Sun Fire 12K, 15K ■ Sun Enterprise 3000, 3500, 4000, 4500, 5000, 5500, 6000, 6500 (PCI and/or sBus, check customer configuration) ■ Sun Enterprise 10000 ■ Sun Enterprise 220R, 250, 420R, 450 (PCI only) ■ 4U architecture: 220, 420, 450, 3000, 3500, 4500, R10 ■ Sunfire models: 3800, 4800, 4810, 6800
SuSE Linux—SLES 8 and United Linux V1.0	■ rx5670, zx6000, rx4640, rx1600
SuSE Linux—SLES 8 (64-bit)	■ ia64

Features

The following are prominent features of the Continuous Access EVA solution:

- In-order synchronous or asynchronous remote replication (remote mirroring)
- Automated failover support
- Support for the following online disk drives:
  - 36-GB 10K and 15K rpm
  - 72-GB 10K and 15K rpm
  - 146-GB 10K rpm
  - 300-GB 10K rpm
- Support for the following near-online disk drives:
  - 250-GB 7200 rpm
- Normal and failsafe data protection modes of operation
- Dual-redundant controller operation for increased fault tolerance
  - No single point of failure within the array
  - Sets of arrays containing two or three members sharing replication relationships
  - Replicated write-back cache support
  - Read-ahead and adaptive read caching support
  - I/O continuation during normalization (full copy) and merging
- Intersite link suspend-and-resume operations
- Multivendor platform support
- Merge of write history log in write order
- Failover scripting
- Multiple-bus failover support
- Continuous Access user interface for management and monitoring
- Asynchronous disk swap (hot swap)
- Controller password protection for configuration control
- Selective storage presentation for SAN-based data zoning

- HP VersaStor-enabled virtual RAID (Vraid) arrays (Vraid1 and Vraid5)
  - Improved performance
  - Improved reliability in Vraid1 and Vraid5
  - Increased disk utilization efficiency

Note: Although Vraid0 is supported, HP does not recommend it in Continuous Access EVA environments.

- Virtual disk data-load leveling
- Clustered server support
- Distributed sparing of disk capacity
- Nondisruptive software upgrade capability
- LiteTouch Management
- Battery back-up
- Bidirectional replication
- Copy set size of 1 GB to 2 TB in 1 GB increments
- Up to 128 remote copy sets
- Up to 128 DR groups
- Up to 8 copy sets per DR group
- Management of up to 512 virtual disks per EVA ranging in size from 1 GB to 2 TB per virtual disk
- Maximum of 240 FC-AL drives (with expansion cabinet) per storage system. Using VCS V3.02 or later the drives can be a combination of online and near-online Fibre Attached Technology Adapted (FATA) drives. The maximum capacity is 35 TB.
- Maximum of 16 storage systems on the same SMA (refer to Table 17 on page 47 for details).
- Maximum of 256 Fibre Channel adapter (FCA) ports per EVA. These ports can be a mix of single and dual channel adapters, that are installed into up to 128 servers (at two FCAs per server) with connections from the FCA to the arrays distributed across the two arrays.

Note: FCA was formerly called HBA or host bus adapter.

- Maximum of 256 LUN presentations from an EVA to a single FCA
- Fabric support at the time of publication. For the most current information, contact your local HP Representative (refer to the *HP StorageWorks SAN Design Reference Guide* for details):
  - Up to 28 B-series switches per fabric; seven hops (interswitch links) are allowed (three per site and one as the intersite link)
  - Up to 16 C-series switches per fabric; seven hops are allowed (where one hop is the intersite link)
  - Up to 24 M-series switches per fabric; three hops are allowed (one per site and one as the intersite link)
- Dark Fibre to 35 km at 2 Gbps or 100 km at 1 Gbps
- 2-Gbps end-to-end Fibre Channel solution
- 100-ms one-way intersite latency with wide area network gateways
- Support for any vendor's wavelength division multiplexing equipment up to the distance that the vendor supports.
- Virtually capacity-free snapshot (Vsnap) function saves significant disk space and improves disk utilization efficiency.
- Virtually instantaneous Snapclone copy capability allows immediate use of the clone copy at the redundancy level of the original volume and can accomplish significant savings of time.
- Snapclone across physical disk groups
- Multiple snaps of the same data (both source and destination)
- Maximum of 8 snapshots or Snapclones per DR group at the local or remote site, or a maximum of 7 per LUN, either source or destination
- The option of a selectable World Wide Name (WWN) for Vsnaps, snapshots, and Snapclones

Other benefits

The following are additional benefits provided by any EVA solution:

- Outstanding self-tuning performance ensures consistency in meeting application Service Level Agreements. Users and clients accomplish more in less time, scale capacity on demand, and minimize data administration overhead.
- State-of-the-art controller software, with VersaStor-enabled virtualization technology, helps improve performance, increases disk utilization efficiency, and allows for easy dynamic storage expansion, all of which help lower costs.
- Supports high-density packaging and more disks per storage array. For example, when using 300-GB drives, achieve up to 35 TB of disk storage in a single cabinet of approximately 5.9 square feet (0.5 square meters).
- Integrated virtualization, enabled by VersaStor technology, improves performance and increases disk utilization efficiency.
- LiteTouch management provides simplified storage management, such as server-independent storage management, automatic load balancing, and instant storage expansion to multiply management efficiency up to 15 times.

Design Trade-offs

2

This chapter discusses the design trade-offs to consider during the review of a Continuous Access EVA solution. This is an informal process, not a set of formal steps. As such, the process is described more in terms of the results you want rather than the specific steps for getting the results.

Topics include:

Business requirements

The first part of the design review process is to understand what business requirements are driving the need for a Continuous Access EVA solution. In some cases, there is a business need for high availability, and disaster tolerance is seen as a side benefit. In other cases, there is a business requirement for disaster-tolerant data, with high availability of the applications being viewed as a side benefit. Neither type of configuration, however, satisfies both requirements without additional hardware or software. Continuous Access EVA provides disaster-tolerant data and supports clustering technology (such as OpenVMS Clusters, TruClusters, MSCS, and VERITAS Clusters) to enable high-availability applications. Together, these solutions provide highly available applications and disaster-tolerant data.

Disaster planning

Your ability to plan for, prevent, and recover from disasters that affect IT operations is critical to maintaining business continuity during a disaster. Disaster planning consists of four broad areas:

- **Business impact assessment**—Includes capacity planning, and risk analysis.
- **Environmental assurance**—Focuses on the physical plant and power requirements for high-availability computing.
- **Disaster tolerance considerations**—Includes protection of environment, infrastructure, data and applications according to downtime requirements.
- **Recovery planning**—Restarts and recovers an IT environment onsite or remotely.

Figure 2 shows how the four solution sets of analyze, design, build/integrate and manage/evolve are interrelated.

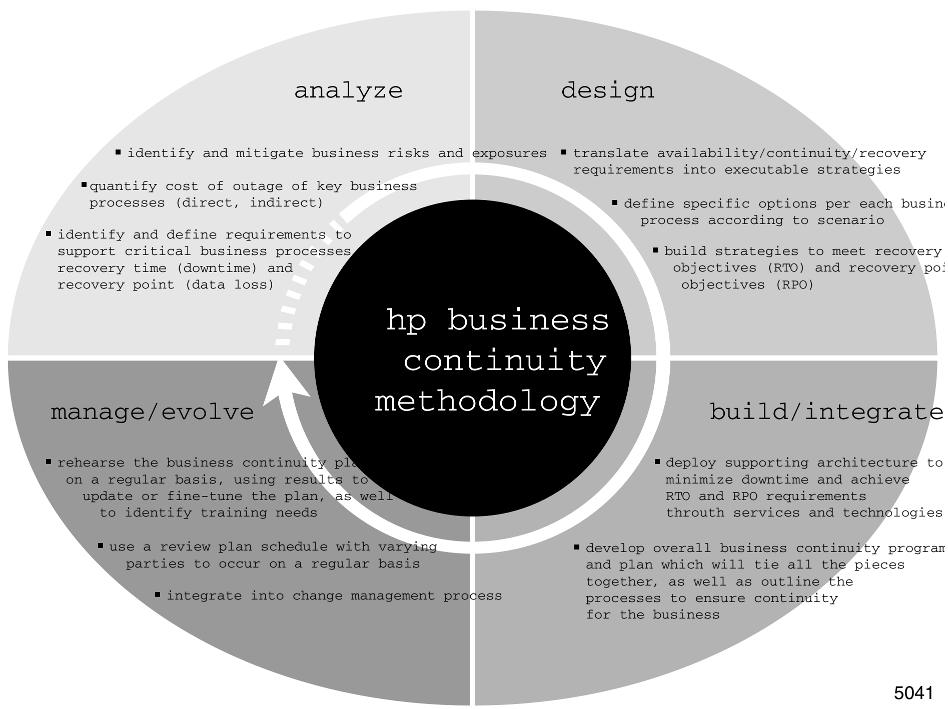

Figure 2: Business continuity model

Continuous Access EVA provides business continuity in the protection and data recovery areas. Because all data is replicated at a remote site, any downtime is minimal. For more information on HP Business Continuity and Recovery Services, browse to http://www.hp.com/hps/tech/continuity/.

High availability

The combination of redundant systems, software, and IT processes, with no single point of failure, reduces the risk of downtime and provides high availability. Continuous Access EVA provides highly available and reliable access to data, but because it is a storage-centric solution, it does not provide highly available applications. To achieve highly available applications, you need standby servers. They provide the processing platforms that the applications can use if the primary server fails.

You can deploy a cluster of servers at the local site, with either a single backup server or cluster of servers at the remote site. The resulting configuration provides highly available and disaster-tolerant applications, and data. Additionally, Continuous Access EVA can function in both directions between sites to optimize the use of the equipment.

Table 4 shows the importance of high availability. Note that even at an availability of 99 percent, an application can be down for nearly four days per year. Those four days could be one long outage or a combined total of several short outages.

Note: The outage includes all time when the application is not available.

Therefore, consider the time it takes to failover the storage, and restart a server and its applications when determining the application Service Level Agreement (SLA).

Table 4: Real-world availability: outage in minutes per year

Availability	90%	95%	99%	99.9%	99.99%	99.999%	100%
Approximate Outage (Minutes per Year)	50,000	25,000	5,000	500	50	5	0

Table 5 illustrates expected user outage minutes per year for several server types.

Table 5: User outage minutes and relative availability

	IBM Parallel Sysplex, HP NonStop Server	Single Mainframe	IBM AS/400, HP OpenVMS, HP3000	Solaris, HP HP-UX, HP Tru64 UNIX, IBM AIX	Windows NT Server & MSCS
Percent Availability	99.99	99.9	99.8	99.6	99.2
Minutes Lost per Year	132	492	1050	2100	4194

Source: Gartner 2001

Disaster tolerance

Disaster tolerance is a special combination of high-availability technology and services that can continue the operation of critical applications in the event of a site disaster. Disaster-tolerant systems are designed to allow applications to continue to operate during the disaster recovery period. For example, if the two halves of a SAN are separated by more than the potential size of a disaster, then one of the sites should be able to continue processing after the disaster. Continuous Access EVA enables applications to automatically and continuously build two copies of application data at geographically separated sites so that a single event will not destroy both sites. However, the greater the distance between the sites, the longer the time required for processing an I/O (as described in the section "Physics of distance" on page 29).

Threat radius

The threat radius is the distance from the center of a threat to the outside perimeter of that threat. For example, half the diameter of a hurricane or typhoon is the threat radius of that storm. On the other hand, consider the release of toxic chemicals from a factory. The shape and size is now defined by the strength of the wind, with the downwind threat radius much larger than the upwind threat radius, producing a more eliptical threat area.

The three general classes of threat in a Continuous Access EVA solution are local, within the metropolitan area, and regional. These three classes make it easier to characterize the solution in terms of intersite link options, performance, and system reliability.

These three types of threats are defined as follows:

- **Local**—Any threat that is less than a few kilometers in radius (less than 25 square kilometers or 15.5 square miles), such as a tornado or fire. Local replication has the least effect on performance compared to the other options.
- **Metropolitan**—Any threat that is larger than a local threat, and which can extend from 25 square kilometers to 5,000 square kilometers (3100 square miles), such as a large chemical incident, a moderate earthquake, or a severe storm. The performance impact due to replication just outside of metropolitan-sized threats is similar in performance cost to running older disk drives—it is slower, but acceptable.
- **Regional**—Any disaster that affects a radius of hundreds of kilometers to tens of thousands of kilometers, such as a large flood or hurricane. By sheer size, the regional disaster requires the largest separation distance when planning disaster-tolerant solutions. Depending on the distances, data replication to the outside of a regional disaster threat radius affects system performance. However, separation distances of more than 1,000 kilometers (620 miles) are rarely needed, and increase the cost of the link and slow down performance rather than provide additional disaster tolerance.

Examples of each disaster include:

- Building fire—Local
- Tornado—Local, but possibly metropolitan if it stays on the ground for a long time
- Hurricane or typhoon—Metropolitan to regional, depending on size and intensity
- Floods—Local to regional along the flood plain
- Earthquake—Usually local to metropolitan, depending on severity
- Environmental—Local to metropolitan, depending on the event
- Power loss—Local to regional, depending on which component fails—a local distribution point or a major portion of the regional power grid

When considering the threat radius, planners must decide what the threats are to the source system, and whether those threats also apply to the backup system. For example, it would be unwise to place both sites in the same flood plain because one flood could destroy both. Similarly, if severe storms tend to travel in a certain direction, then a good strategy would be to place the second site perpendicular to the expected route of travel, and as far apart as needed to prevent one storm from affecting both sites.

Figure 3 on page 28 is an example of the relative relationship between the three classes of threats based on the radius of those threats.

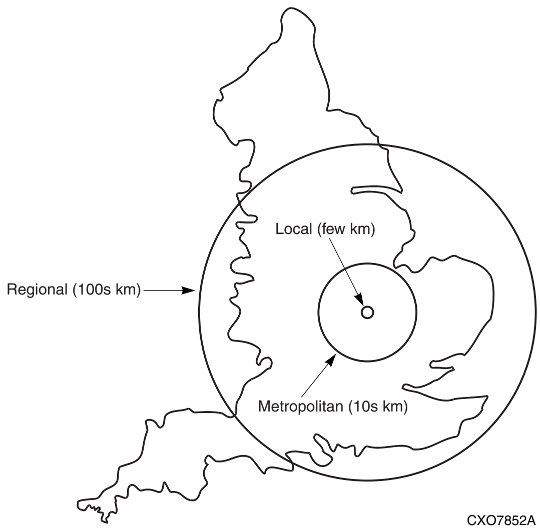

Figure 3: Threat radius

The larger the potential disaster (threat radius), the farther apart the local and remote sites should be. To determine an adequate distance, determine what kinds of disasters are probable in the local area and understand the protection distances needed to separate the two sites. Consider any local regulatory requirements that can increase or limit the separation distance. For example, some counties in the United States require both sites to remain within the same 100- to 400-square-kilometer (60- to 250-square-mile) county. This restriction has limited the maximum separation distance to less than 30 km (19 mi) in an area prone to earthquakes. Such earthquakes have affected buildings several hundred kilometers from the earthquake epicenter.

As another example, on the East Coast of the United States, and the southern and eastern coasts of the Asian subcontinent, hurricanes or typhoons can cover an area with a radius of 200 km (125 mi) or more. Other natural disasters include regional forest fires, localized tornadoes, and widespread flooding. Unnatural disasters include building fires or chemical contamination, either of which can limit access to computer facilities. In these cases, the threat can move with respect to the fixed facilities. It is, therefore, necessary to ensure that the two sites are sufficiently far apart, so that the same disaster does not affect both sites.

Physics of distance

To deal with the types and scope of potential disasters, storage system planners must consider the inherent trade-off: the greater the distance between the two sites, the lower the maximum performance that is possible. This is due to the increased time it takes to complete a replication as separation distances are also increased.

One-way delays

The speed of light in a vacuum, for example, is about 3 x 10^8 meters per second or 186,000 miles per second. In fiber optic cables, this slows down to about two-thirds of the speed of light in a vacuum, or approximately 2 x 10^8 meters per second. Converting the speed of light in fiber from meters per second to seconds per meter, the result is 5 nanoseconds per meter, or 5 microseconds per kilometer. This means that it takes time for the bits in a read or a write to move from one place to another. In a larger context, at the time this version of the document was published, the maximum Continuous Access EVA separation is limited to a one-way delay of 100 milliseconds. This is equivalent to a cable distance of 7200 kilometers, or roughly 4500 miles. Other examples of one-way delays are shown in Table 6.

Table 6: Examples of one-way delays

One-Way Delay (ms)	Point-to-Point Cable Distance (km)	Point-to-Point Cable Distance (miles)
1	200	125
3	600	375
9	1800	1125
18	3600	2250
36	7200	4500
60	12,000	7500
100	20,000	12,500
Note: For both kilometers and miles, routed networks will be shorter in cable distances due to the routing delays.		

Effects of distance on sustainable I/O rates

All HP Continuous Access products can move data at extreme distances, but due to propagation delays inherent in data transfers, it can take a long time for each replication I/O to complete. The long distance between the two sites, not the bandwidth of the communications link, is usually the limiting factor in replication performance.

Figure 4 shows two I/O links: a thinner arrow representing a lower bandwidth communications link, and a wider arrow representing a higher bandwidth intersite link. Both I/O packets (shown by the parallelograms) contain the same amount of data. They are moving from Site A to Site B (left to right), and the leading edge of both will arrive at Site B at the same time. The difference in link size (bandwidth of the link) allows the wider link to complete delivery of the packet before the narrower link. The wider link provides shorter times to load and unload the packet. In terms of something visible, a nanosecond of fiber is roughly 8 inches long and, at 2 Gbps, contains just 2 bits of data. At 1 Gbps, a single bit is 8 inches long. At 100 Mbps, a bit is 80 inches long and at 10 Mbps the single bit is 800 inches long.

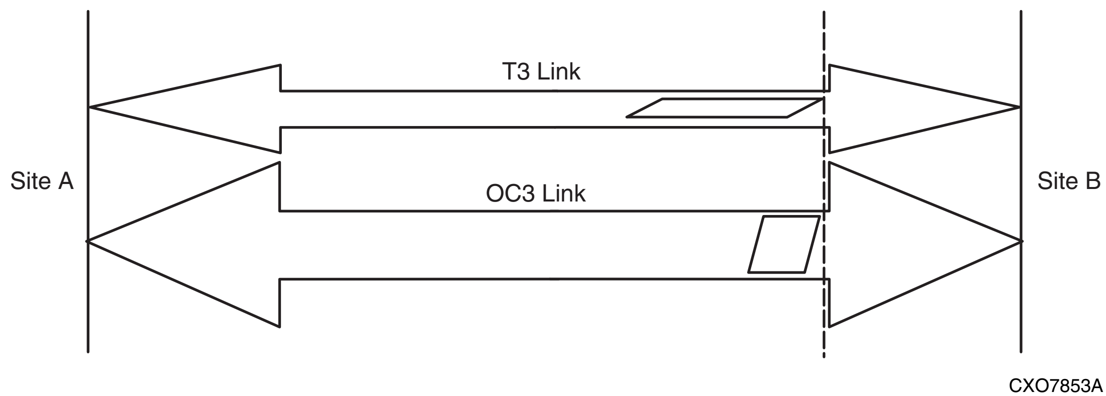

Figure 4: Effect of distance on I/O rate

Given the bandwidth of various links, it is possible to calculate how long it takes for a given amount of data to be inserted into and removed from the SAN, and is called the *time to load* the data. For example, given a 16-KB packet and a T1 transmission speed of 1.54 Mbps, it takes approximately 0.1 second to load the packet onto the link. Using a Fibre Channel link running at 2 Gbps and the same 16-KB data packet, the time is reduced to 0.00043 seconds.

Knowing both the effect of distance and bandwidth, Table 7 on page 31 provides the data needed for calculating how long it takes to complete a single I/O replication across a zero distance intersite link, based on the link bandwidth and the size of the data packet. The columns are the transmission technology, the bandwidth of that technology, with the Y-intercept *b* representing the replication and conversion overhead for a single 512-byte packet, and the slope *m* representing the additional time needed for transfers larger than 512 bytes.

The following formula shows how to calculate *Y* (the ideal time to complete a single replication I/O) given *X* as the size of the data packet in kilobytes, and using *m* and *b* from Table 7:

$$Y \text{ (in milliseconds)} = mX + b$$

This data was obtained experimentally and is an average for a particular technology.

Table 7: Link performance table

Intersite Link Technology	Intersite Link Bandwidth (Mbps)	Intercept b	Slope m
2-Gbps Fibre Channel	2000	0.3416	0.0268
1-Gbps Fibre Channel	1000	0.3991	0.0332
1-GbE (Gigabit Ethernet) IP	1000	0.4130	0.0338
OC3 IP	155.5	0.3901	0.0758
E4 IP	139.3	0.3876	0.0818
100-Mbps IP	100	0.3802	0.1052
T3 IP	44	0.3530	0.2070
E3 IP	34.304	0.3340	0.2666
10-Mbps IP	10	0.2893	0.8872
E1 IP	2.048	1.1748	4.4434
T1 IP	1.54	1.5557	5.9422

Note: (1) b = the replication and conversion overhead—the amount of time required to replicate one 512-byte block.

(2) m = a number proportional to the additional amount of time it takes to load and unload larger packets versus smaller packets.

(3) These values are derived experimentally. There is a nonlinear relationship between the bandwidth and the slope of the line, and between the bandwidth and the intercept of the line with the y-axis.

To calculate how long it takes to complete a replicating I/O over distance, add the effects of distance to the time it takes to complete a zero distance replication I/O. The effects of distance can be determined by calculating the time it takes for the leading edge of the data to travel from one end of the link to the other. Estimate this time using a speed of 2×10^8 meters per second for light in standard fiber-optic cable, which is equal to 5 microseconds per kilometer. For example, an intersite separation distance of 10,000 km (6,200 mi), the one-way time interval is 50,000 microseconds, or 0.05 seconds. Note that this is about 25 times the average rotational latency of a 15,000-rpm disk drive at 0.002 seconds.

Before proceeding with the calculation, note that Continuous Access EVA requires one round-trip through the intersite link to complete a single replication. First, the data is sent from the source array to the destination array. When the destination controller receives and stores the data in its local cache, it then returns an acknowledgement back to the source controller. Therefore, the true time it takes to move data from the source array to the destination array is the single round-trip distance, plus the time it takes to load and unload the data for a given link size. These trips, consisting of one small control packet and one large data packet, add 10 microseconds of latency per kilometer of one-way intersite link distance to complete each replication write. Based on the distance between the two sites, this latency is added to the previously calculated time to complete a zero distance replication write.

Intersite latency

Two methods are used to determine the intersite latency based on the cable distance between the two sites. One is based on the network latency, the other on the driving distance. If an intersite network exists, use the more accurate network latency as the metric. If an intersite network does not exist, then use the driving distance approach. In either case, allow an additional 50 percent latency if you are converting a point-to-point dedicated network to a shared, routed network.

Network latency

If an intersite network connection exists, ask the network engineers for an estimate of the one-way or round-trip intersite latency. For example, consider two sites that are separated by a straight line distance of 150 km (93 mi), yet the actual point-to-point network was measured to give one-way latency of approximately 1 ms or 200 km (125 mi).

Note: The Internet protocol "ping" utility usually reports the round-trip latency.

Driving distance

To estimate the cable distance between two sites without a network, measure the distance by driving a car between the two sites. For point-to-point networks, multiply the driving distance by 1.5. For routed networks, multiply the driving distance by 2.25 to account for the additional delays due to network routing. Using the preceding example, the estimated cable distance would be 225 km (140 mi) for a point-to-point network instead of the 200 km (125 mi) based on the actual latency. For improved estimation, contact one or more possible network vendors and ask about possible routing and delays between the two sites in question. If a delay is known use it; if not, estimate the driving distance along the network path, and not the shortest distance between the two points.

Higher than expected latency

If you can, compare the results of both methods. If the current network latency is more than five times the number determined by the driving distance, consider using alternative networks with a lower latency to link the two sites.

Another source of higher than expected latency can be over subscription of the existing network. To check for over subscription, look for the packet loss ratio. Continuous Access EVA does not support a packet loss ratio greater than 0.2 percent after the addition of Continuous Access EVA traffic.

HP recommends that any network which supports a storage interconnect or remote replication not have a packet loss ratio that exceeds an average of 0.01 percent over 24 hours.

Single-stream I/O example

For the purpose of this example, consider an application that can perform only one write at a time and must wait for that write to conclude before the next write is issued. This type of application is defined as a *single server-based, synchronous I/O stream*. Further assume that the replication is also set to synchronous, meaning that the write must be in both the source and destination controllers' cache before the write is acknowledged as complete back to the issuing server.

This example assumes that each of these writes consists of a 32-KB block of data, and that the local and remote sites are located 200 km (125 mi) apart, based on the one-way network latency of 1 ms.

Using the packet size, Table 8 shows the results of the time to load the 32-KB packets onto various networks.

Table 8: Calculation of time to load data onto an intersite link

Technology	Slope	Time to Load 32 KB of Data (ms)
2-Gbps Fibre Channel	0.0268	0.86
1-Gbps Fibre Channel	0.0332	1.06
1-Gbe (Gigabit Ethernet) IP	0.0338	1.08
OC3 IP	0.0758	2.43
E4 IP	0.0818	2.62
100-Mbps IP	0.1052	3.37
T3 IP	0.2070	6.62
E3 IP	0.2666	8.53
10-Mbps IP	0.8872	28.39
E1 IP	4.4434	142.19
T1 IP	5.9422	190.15

Given the intersite distance of 200 km, the transfer (round trip) latency is:

$$200 \text{ km} \times 5\ \mu\text{s/km} \times 2 = 2 \text{ ms}$$

and is presumed to be the same for all types of interconnects.

Table 9 provides the calculation for adding the results in Table 8, plus the latency and the replication overhead.

Table 9: Calculating time to complete a single 32-KB write I/O over distance

Technology	Time to Load (ms)	+ Transfer (round-trip) Latency (ms)	+ Overhead (ms)	= Time to Complete I/O (ms)
2-Gbps Fibre Channel	0.86	2	0.3416	3.20
1-Gbps Fibre Channel	1.06	2	0.3991	3.46
1-GbE (Gigabit Ethernet) IP	1.08	2	0.4130	3.49
OC3 IP	2.43	2	0.3901	4.82
E4 IP	2.62	2	0.3876	5.01
100-Mbps IP	3.37	2	0.3802	5.75
T3 IP	6.62	2	0.3530	8.98
E3 IP	8.53	2	0.3340	10.87
10-Mbps IP	28.39	2	0.2893	30.68
E1 IP	142.19	2	1.1748	145.36
T1 IP	190.15	2	1.5557	193.71

Inverting the time-to-complete number in milliseconds per I/O produces the maximum number of single-stream synchronous replication writes in I/Os per second (IOPS) that can be completed every second for the various bandwidths of links for the example distance. This assumes that the next I/O is started immediately upon completion of the previous I/O.

The results of the inversion are shown in Table 10 as the approximate single stream I/O rates and throughput based on this simple example. Because it is a simple case, and most applications produce multiple asynchronous I/Os, the result is the minimum expected performance. In this table, throughput is calculated as the packet size times the I/O rate, and percentage of bandwidth used is the peak bandwidth of the link, divided by the throughput, times 100.

Table 10: Relationship of I/O per second to throughput and percent bandwidth used

Technology	Approximate IOPS	Throughput (Mbps) based on 32KB per I/O	Approximate Single Stream Percent Bandwidth Used
2-Gbps Fibre Channel	312.58	100.03	5.0%
1-Gbps Fibre Channel	288.89	92.45	9.2%
1-GbE (Gigabit Ethernet) IP	286.16	91.57	9.2%

Table 10: Relationship of I/O per second to throughput and percent bandwidth used

Technology	Approximate IOPS	Throughput (Mbps) based on 32KB per I/O	Approximate Single Stream Percent Bandwidth Used
OC3 IP	207.65	66.45	42.9%
E4 IP	199.79	63.93	46.0%
100-Mbps IP	174.02	55.69	55.7%
T3 IP	111.40	35.65	79.2%
E3 IP	92.04	29.45	86.6%
10-Mbps IP	32.59	10.43	100.0%
E1 IP	6.88	2.20	100.0%
T1 IP	5.16	1.65	100.0%

You can also use the *HP StorageWorks CA EVA Replication Performance Estimator* to perform these same calculations. This tool, along with Continuous Access EVA documentation, is available on the HP website at http://h18006.www1.hp.com/products/storage/software/conaccesseva/index.html. Select "related information" from the list on the right side.

Multiple I/O streams for a single application

Most applications do not issue a single I/O and then wait for it to complete before immediately sending the next one. Rather than waiting, they have multiple I/Os open at any one time (called asynchronous server I/O).

After determining the worst case time needed to perform the replication of a single write, you must examine the effect of multiple I/O streams from a single application. Multiple writes can be replicated at the same time, as shown in Figure 5 on page 37. The wider the link, the more writes it can hold, up to a maximum called the bandwidth-latency product. Multiply the net performance of the communications link (total bits per second, less any overhead) by the one-way intersite latency in seconds. This number, when expressed in bytes (Fibre Channel uses 10 bits per byte) and divided by the average message size (in bytes), determines the absolute maximum number of messages that can be in the communications link at one time.

At this point in the process a basic assumption exists—that the application is able to issue an I/O as soon as the link is open, thus streaming the data from the server to the first EVA. This data is then replicated to the destination EVA with acknowledgment back to the server. A further assumption exists—that because there is only one source, the data is written fairly regularly and consistently. Both assumptions are necessary because the purpose of this example is to understand peak performance needs.

In Figure 5 on page 37 the parallelograms represent multiple-write data streams from the same application. Each I/O stream consists of data packets of equal size (same number of bytes). The narrow link is able to hold only three packets, but the wider link is able to hold nine packets. This example assumes that each new packet immediately follows the preceding packet regardless of link bandwidth.

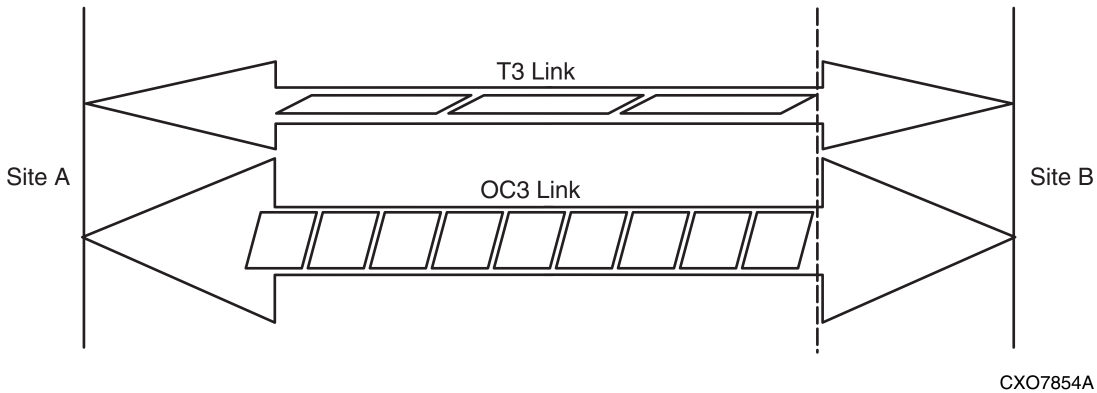

Figure 5: Effect of streaming multiple I/Os for a single application

Using the data from Table 10 on page 35, consider the values in the last column—the percentage of bandwidth used by the single I/O stream. Assuming there is minimal separation between any two writes, inverting the percentage of bandwidth number and dividing by 100 results in the approximate peak number of I/O streams that can be supported by that link for that distance. Multiplying this peak number of streams times the single I/O rate produces the theoretical peak write rate for the application.

However, four constraints limit the expected maximum number of I/O streams in the link, which becomes an upper boundary on the maximum performance that can be expected for a given set of distance, packet size, and bandwidth.

The first limit is the raw capability of the server, FCA, and source HSV storage controller and array of attached disks. For example, to ensure write order across all members of a DR group, the members are restricted to using the same FCA port and EVA controller and load balancing is turned off.

The second limit is the number of Fibre Channel buffer-to-buffer credits allowed between any two Fibre Channel devices. For B-series switches, the credits are specified in Table 11. For C-series switches, the credits are specified in Table 12. For M-series switches, the credits are specified in Table 13 on page 39. This limit becomes the bottleneck on any long-distance direct Fibre Channel connection with very long-distance gigabit interface converters (GBICs) or a wavelength division multiplexing (WDM) solution. It is not usually seen in Continuous Access EVA over IP because the credit is returned by the IP gateway to the sending switch. Note that the B-series switch limit can be increased up to 60 when using the Extended Fabric License.

Table 11: B-series switch FC buffer-to-buffer credits

Switch Family	Default Credits	Credits Available with Extended Fabric License
3xxx at 1 Gbps	16	60
4xxx at 1 Gbps	27	60
3xxx at 2 Gbps	27	64
4xxx at 2 Gbps	27	64

Table 12: C-series switch FC buffer-to-buffer credits

Switch Family	Default Credits	Maximum Credits Available
All 16-port modules, Fx mode	16	255
E or TE modes	255	255
All 32-port modules	12	N/A

Table 13: M-series switch FC buffer-to-buffer credits

Switch Family	Default Credits	Maximum Credits Available
3xxx	60	60
4xxx	162	162
6xxx	60	60

The third limit is the maximum number of outstanding replication I/Os the HSV controller can support. Using VCS versions 3.00 or 3.01, the HSV controller allows up to 32 (8 KB) outstanding replication I/Os per controller port. If the bandwidth latency product of the link can support more than 32 I/Os, such as in high speed, very long distance Continuous Access EVA-over-IP configurations, the maximum number of 8 KB or smaller I/Os outstanding in the link will be 32 per HSV controller link sharing the link.

In an environment where the controllers are using VCS V3.02 or later, up to 64 outstanding replication I/Os are allowed per controller port. In a worst-case situation with one port of each controller using the same link, the combined limit is 128 (64 for each port) outstanding I/Os, subject to other limits.

The fourth limit is the impact of the initial full copy process on the intersite link. In all V3 versions of VCS, the full copy process allocates eight 128 KB data buffers. Replication consumes as much as 64 x 8 KB (or 512 KB), but the full copy consumes as much as 8 x 128 KB (or 1 MB). These processes combined can overwhelm an intersite link if it has too little bandwidth.

Multiple I/O streams for multiple applications

The environment shown in Figure 6 on page 40 is similar to situations where multiple I/O streams are issued by multiple applications. The primary difference between Figure 6 and the single application environment in Figure 5 on page 37 is that with multiple applications (across one or more servers), the space between the I/Os is wider than with the single application model. This tends to reduce the number of writes in the intersite link and also reduces the maximum utilization rate of that link. Any differences in the number of I/Os in the link are based on how the I/O bandwidth is shared between the competing applications.

If there are multiple high-performance applications, HP recommends adding more Continuous Access EVA storage arrays to maintain performance.

The average loading on any link must not exceed 40 percent of its rated capacity, and the peak loading must not exceed 45 percent of rated capacity. This limitation allows an I/O from a failed link or fabric to run on the nonfailed fabric or link without causing additional failures due to overloading of the surviving fabric.

Note that in most real-world application environments it is not possible for all writes to arrive at precisely the correct time, such that there is minimal space between any particular write, the one before it, and the one after it. In other words, it is not possible to fill the pipeline end-to-end with data. For example, when 10-Mbps Ethernet first became available, it was believed that the maximum effective throughput was 10 Mbps. Instead, the theoretical maximum is approximately 70 percent of the 10 Mbps figure, and in practice a maximum utilization figure of approximately 50 percent is seldom achieved.

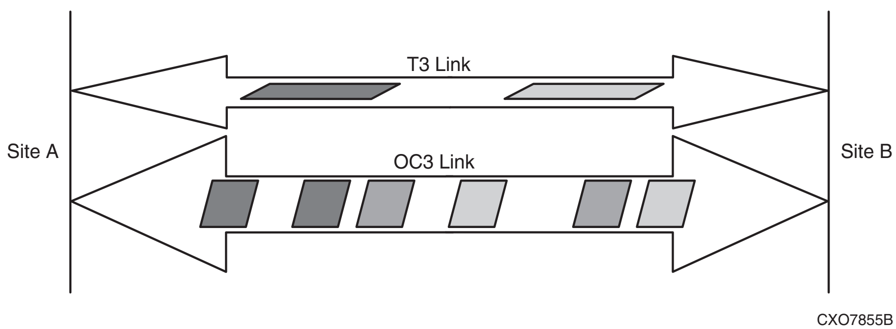

Figure 6: Effect of streaming multiple I/Os for multiple applications

Multiple I/O streams example

This example illustrates how to predict the effect of multiple independent writes on the intersite link. Assume there are two different applications—one performs large writes of 32 KB and the other performs small writes of 2 KB. Because the large write is the same size as the example shown in Table 10 on page 35 that data is used here for the large I/O stream. To keep it simple, the same steps are repeated for 2 KB writes. The results are shown in Table 14.

Table 14: Relationship of I/O per second to throughput & percent bandwidth used for 2-KB writes

Technology	Approximate IOPS	Throughput (Mbps) based on 2 KB per I/O	Approximate Single Stream Percent Bandwidth Used
2-Gbps Fibre Channel	417.50	8.35	0.4%
1-Gbps Fibre Channel	405.60	8.11	0.8%
1-GbE (Gigabit Ethernet) IP	403.13	8.06	0.8%
OC3 IP	393.44	7.87	5.1%
E4 IP	391.97	7.84	5.6%
100-Mbps IP	386.01	7.72	7.7%
T3 IP	361.40	7.23	16.1%
E3 IP	348.77	6.98	20.5%
10-Mbps IP	246.08	4.92	49.2%
E1 IP	82.91	1.66	82.9%
T1 IP	64.77	1.30	86.4%

The next step is to determine the actual number of IOPS expected from each of the two applications. This is done by scaling the single I/O rate either up or down to the anticipated rate, and then scaling the bandwidth needed by the same number. If at any time the total required bandwidth exceeds 25 percent, that link should not be expected to support the required communications and should be dropped from further consideration. For this example, one application read and write produces transactions at 3 times the 32-KB single-stream rates, and the other read and write produces transactions at 4 times the 2-KB single-stream rates. The resulting intermediate numbers are shown in Table 15.

Table 15: Example summary data for multiple applications writing across the same link

Technology	32 KB IOPS times 3	32 KB Throughput (Mbps)	Percent Bandwidth Used	2 KB IOPS times 4	2 KB Throughput (Mbps)	Percent Bandwidth Used	Percent Total Bandwidth Required
2-Gbps Fibre Channel	938	300	15%	1670	33	2%	17%
1-Gbps Fibre Channel	867	277	28%	1623	32	3%	31%
1-GbE (Gigabit Ethernet) IP	859	274	27%	1613	32	3%	31%

Table 15: Example summary data for multiple applications writing across the same link (continued)

Technology	32 KB IOPS times 3	32 KB Throughput (Mbps)	Percent Bandwidth Used	2 KB IOPS times 4	2 KB Throughput (Mbps)	Percent Bandwidth Used	Percent Total Bandwidth Required
OC3 IP	623	199	129%	1574	31	20%	149%
E4 IP	599	192	138%	1568	31	23%	161%
100-Mbps IP	522	167	167%	1544	31	31%	198%
T3 IP	334	107	238%	1446	29	64%	302.%
E3 IP	276	88	260%	1395	28	82%	342%
10-Mbps IP	98	31	313%	984	20	197%	510%
E1 IP	21	6.6	330%	332	6.6	332%	662%
T1 IP	15	5	330%	259	5.2	345%	676%

As Table 15 shows, for this example only, the 2-Gbps link is able to support the replication requirements. If high-speed links are not available, multiple low-speed links can be used to obtain the required 25 percent of the available bandwidth.

A quick review of the 25 percent rule is in order. Independent applications do not coordinate with each other when sending reads or writes to the source EVA. This lack of coordination causes a space to appear between any two writes on the intersite link. Mathematical queue theory suggests that the expected peak utilization will be only 70 percent, with the expected average around 50 percent. In addition, there must be sufficient bandwidth on both fabrics for all of the traffic, in case one fabric fails. Assuming there is an even split between the two fabrics, the 50 percent becomes 25 percent for planning purposes only.

Caution: Do not design a solution that will overload a controller during normal operations. This prevents significant loss of performance and possible loss of replication data during failure conditions due to the link going down. This is especially true when using asynchronous replication and during the initial full copy or normalization process that happens when a new DR group is created.

Performance impact of multiple replication relationships

The section “Multiple I/O streams for a single application” on page 36 discussed the following three constraints, which limit the expected maximum number of I/O streams in the link:

- The capability of the server, FCA, and source HSV storage controller and array of attached disks.
- The number of Fibre Channel buffer-to-buffer credits allowed between any two Fibre Channel devices.
- The maximum number of outstanding replication I/Os the HSV controller can support.

In the case of multiple relationships, only the third constraint applies to each relationship; the other two apply to the whole solution. In other words, for each replication relationship, there is a set of buffers in the storage system that are used to process the replication I/Os. In Figure 7 the storage system in Site A contains two sets of buffers, one set of buffers for the relationship between Sites A and B, and one set of buffers for the relationship between Sites A and C.

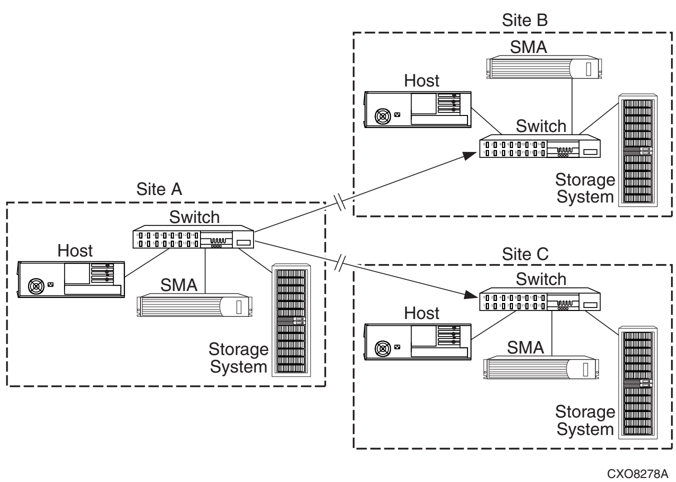

Figure 7: Buffers in storage systems

The other two constraints (capability of the array and buffer-to-buffer credits) apply to both relationships. Therefore, in some cases it may be possible to get more aggregate performance out of two pairs of arrays, each sharing the two destinations, rather than using two single relationship replicating pairs with each having a dedicated destination. Compare Figure 8 and Figure 9 on page 45.

In Figure 8, both Sites A and D share destination Sites B and C, which in some solutions can outperform the example in Figure 9.

Note: While this solution optimizes the performance, it requires careful planning as each link supports different DR groups.

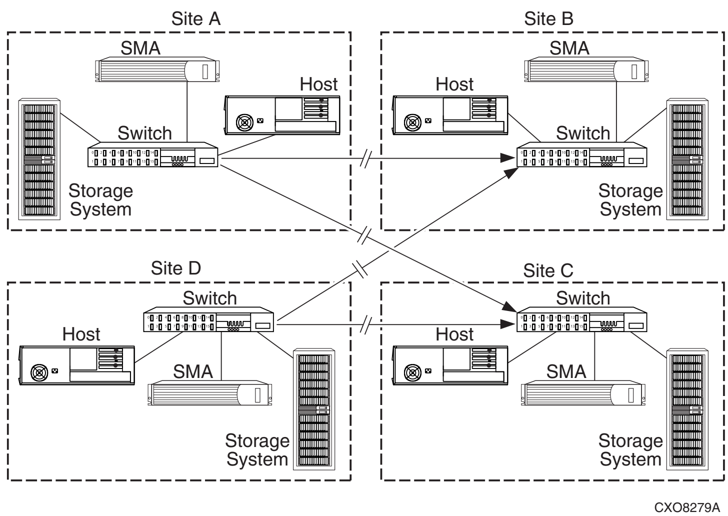

Figure 8: Four relationships sharing destinations

In Figure 9, Sites A and D have dedicated destinations, Sites B and C, respectively.

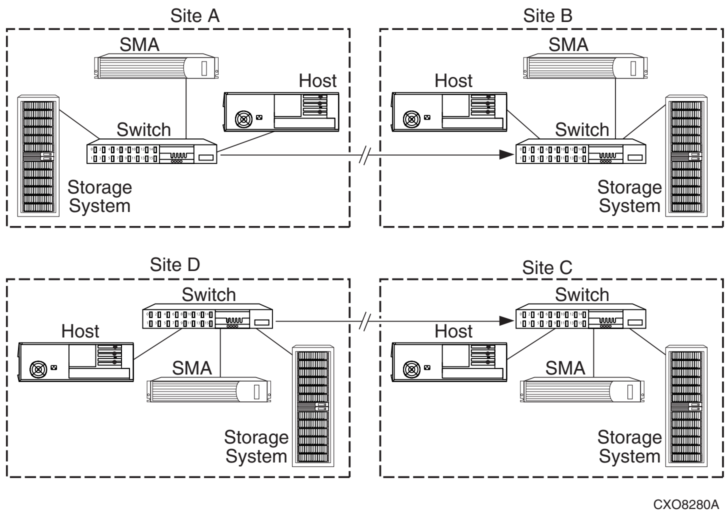

Figure 9: Two relationships with dedicated destinations

It is the beyond the scope of this guide to give exact guidance in this design decision. Empirical evidence shows that the effect of using multiple relationships to improve performance occurs whenever the overall performance is limited by distance, if there is sufficient bandwidth in the links and capability in the array, FCA, and servers.

As an example of how you can use multiple relationships, consider the example shown in Tables 8, 9, 10, and 14 and then summarized in Table 15 on page 41. Suppose that instead of combining the two I/O streams onto one link, a multi-relationship fan-out solution were proposed similar to Figure 7. For this new example, allow the 3 times 32 KB I/O replication stream from Tables 8, 9, and 10 to use one of the two links, and the 4 times 2 KB I/O replication stream from Table 14 to use the other link. Table 16 (extracted from Table 15), shows the 32-KB example workload would require 27 percent of a 1 GbE link, and the 2-K example workload would require 31 percent of a 100 Mbps Ethernet link. Both of these workloads are supportable on their separate links.

Table 16: Example of multiple relationship fan out solution

Technology	32 KB IOPS times 3	32 KB Throughput (Mbps)	Percent Bandwidth Used	2 KB IOPS times 4	2 KB Throughput (Mbps)	Percent Bandwidth Used
1-GbE (Gigabit Ethernet) IP	859	274	27%	1613	32	3%
100-Mbps IP	522	167	167%	1544	31	31%

Determining write-to-disk rate

Given peak write requirements for existing applications, the basic question—whether a proposed solution will work—can now be answered.

For existing applications, determine the current I/O performance rates for each logical unit (LUN) that is to be replicated to the remote site. Finding the average and maximum write rate (write I/O per second) and peak write transfer rate (bytes per second) can require some analysis and time. Use operating system-specific tools, such as *PERFMON* or *IOSTAT*, to collect the data. If the data is only available as an average over time, attempt to ascertain the peak hours of operation and estimate the peak write rate and write transfer rates.

Record the peak and average write rates, and write transfer rates and, if bidirectional, record these numbers for each direction. Compare these numbers to potential intersite links to see if the link technology is affordable. If not, look at other ways to replicate the data, or replicate only the most critical data, such as transaction or redo logs.

The average loading on any link must not exceed 40 percent of rated capacity, and the peak loading must not exceed 45 percent of rated capacity. This limitation allows I/O from a failed link or fabric to run on the nonfailed fabric or link, without causing additional failures of the surviving fabric.

One final consideration is that as distance (or intersite latency) increases, the distance limits the maximum I/O rate more than the bandwidth. For example, if the calculations used in the previous examples were repeated using 100 msec of one-way latency instead of 1msec, regardless of the bandwidth, the math shows only five writes per second are possible for a single replication stream.

Other considerations

This section discusses design trade-offs as they relate to multiple EVA pairs versus distance, booting, and trunking.

EVA management versus distance

This section describes the effect of distance on the management of EVAs that are not local to the active Storage Management Appliance. In addition to distance, the size of the configuration (the number of objects) to be managed also affects the time it takes to perform some of these management actions. For the purpose of estimating the size of a configuration, consider each disk group, disk drive, defined server, virtual disk, DR group, and copy set to be an object that must be managed. The more objects there are, the longer it takes to discover them and manage the array. Similarly, the more remote the array, the longer it takes. To manage a single object, combining a configuration with many objects and extreme distances can require more time to manage than is acceptable.

Table 17 on page 47 is based on the observation that the practical time limit to wait for a management action (that is, discovery of what is in an array) to complete is about 10 minutes. If you allow more time to discover an array, then more arrays could be managed at greater distances. Similarly, it can take roughly as much time to discover four small configurations as it does to discover two large configurations. Conversely, if the allowable discovery time is only five minutes, then you should plan to manage fewer arrays for any given distance listed in Table 17.

Table 17: Distance versus array manageability

Legend: ✓ = supported, recommended; — = not recommended; ns = not supported

Number of Remote EVAs and Relative Config Size	Campus (less than 10 km)	Metro (out to 200 km or 1 msec)[1]	Regional (1 msec to 18 msec)[1]	Multiple Regions (18 msec to 36 msec)[1]	Intra-continental (36 msec to 60 msec)[1]	Inter-continental (60 msec to 100 msec)[1]	Global (greater than 100 msec)[1]
1 small[2]	✓	✓	✓	✓	✓	✓	ns
1 large[3]	✓	✓	✓	✓	✓	—	ns
2 small[2]	✓	✓	✓	✓	✓	—	ns
2 large[3]	✓	✓	✓	✓	—	—	ns
4 small[2]	✓	✓	✓	✓	—	—	ns
4 large[3]	✓	✓	—	—	—	—	ns

Table 17: Distance versus array manageability (continued)

Legend: ✓ = supported, recommended; — = not recommended; ns = not supported

Number of Remote EVAs and Relative Config Size	Campus (less than 10 km)	Metro (out to 200 km or 1 msec)[1]	Regional (1 msec to 18 msec)[1]	Multiple Regions (18 msec to 36 msec)[1]	Intra-continental (36 msec to 60 msec)[1]	Inter-continental (60 msec to 100 msec)[1]	Global (greater than 100 msec)[1]
8 small[2]	✓	✓	—	—	—	—	ns
8 large[3]	✓	—	—	—	—	—	ns

[1] These are one-way latencies.

[2] A small array configuration consists of 1 server using 3 DR groups and 2 copy sets per DR group, for 6 virtual disks built out of 60 disk drives in one disk group.

[3] A large array configuration consists of 10 servers using 64 DR groups and 64 copy sets, for 64 virtual disks built out of one disk group of 24 disks.

Trunking, port channels, or open trunking

New generation B-series switches operating at 2 Gbps offer a feature called *trunking*. C-series switches offer a similar capability called *port channel*. Trunking is the combining of two or more low-speed intersite links (ISLs) into one virtual high-speed ISL to increase link performance. Trunking is enabled by the switches when more than one ISL exists. The effect of trunking is to create a wider link between the two switches that is capable of supporting more transactions. It does not reduce the interswitch latency based on the separation distance.

For M-series switches, this capability is called *open trunking*. Refer to the *HP StorageWorks Continuous Access EVA V1.1B Release Notes* for support statements, and the vendor documentation for additional details about these features.

Replication protocol

This section describes the basic Continuous Access EVA replication protocol.

Controller command and I/O protocol

Within the EVA, the HSV controller uses a specially developed, high-performance protocol for peer-to-peer remote replication I/O. Each of these replication I/O messages includes a unique sequence number to ensure that remote writes are not reordered. On arrival at the destination controller, the arrival sequence is verified and the remote writes are applied to the remote copy in the same order as in the local copy. The controllers also configure the switches to which they are attached for in-order delivery. If there is a problem in the fabric that prevents the in-order delivery of a set of messages, such as a single message that does not get delivered, the sending controller resends the missing message. As a result, whether using either synchronous or asynchronous replication, the destination always contains "crash-consistent" data, although copies can lag behind the source copy of the data.

Likewise, the controller automatically handles performance fairness to virtual disks not configured for data replication on the same controller. The protocol ensures that the data-replication virtual disks do not occupy all the resources. Other optimizations include examining new incoming writes that occur while a full copy is in progress, to ensure that they are not replicated if the block range of the write is within the area that has yet to be copied.

Synchronous versus asynchronous replication protocol

Traditionally, you can choose from two modes of operation for remote replication of data: synchronous and asynchronous. For long distances, synchronous replication mode can impose unacceptable latencies (response time) in returning write completion to the server. Asynchronous mode can reduce the response time (write completion back to the server), but it puts each of the outstanding writes at risk if the source storage system is lost between the time the write is acknowledged as complete to the server and it is written to the destination controller.

When in synchronous mode, the write proceeds as follows (see Figure 10):

1. Data is received from the server and stored in the source controller cache.
2. Data is sent from the source controller to the destination controller cache.

3. When the destination controller receives and stores the data, it returns an acknowledgement to the source controller.
4. When the source controller receives the acknowledgement from the destination controller, it returns an I/O completion acknowledgement to the server.

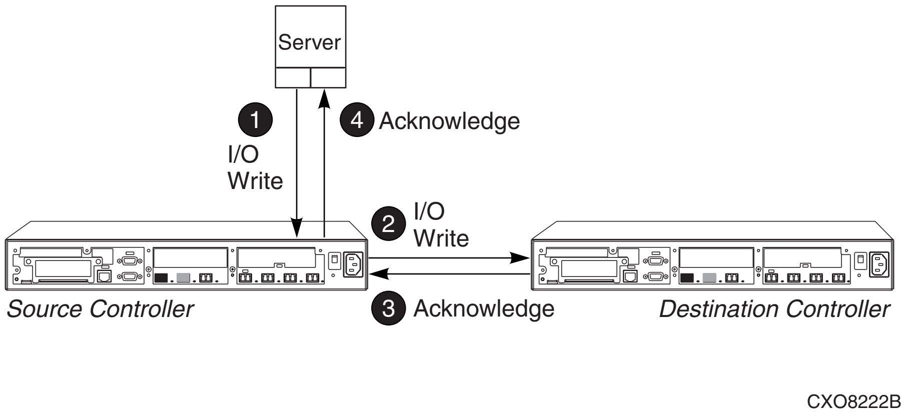

Figure 10: Synchronous replication sequence

When in asynchronous mode, the write proceeds as follows (see Figure 11 on page 51):

1. Data is received from the server and stored in the source controller cache.
2. Completion of the I/O is sent to the server by the source controller.
3. Data is sent from the source controller to the destination controller cache.
4. When the destination controller receives and stores the data in cache, it returns an acknowledgement to the source controller.

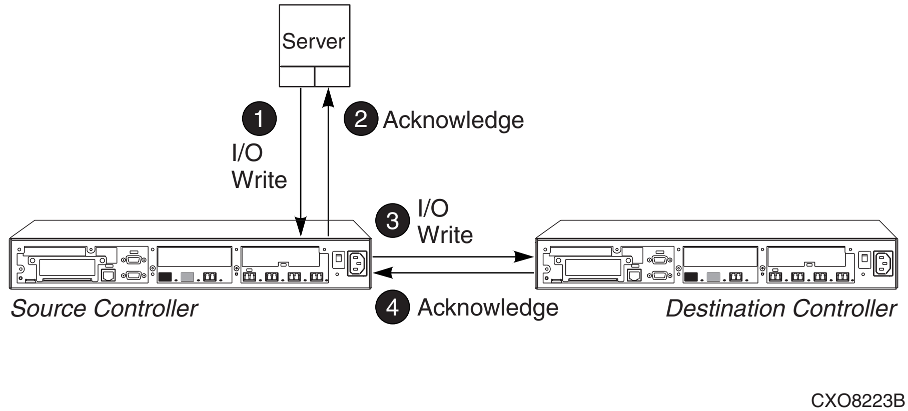

Figure 11: Asynchronous replication sequence

Figure 12 shows the relative performance differences between synchronous and asynchronous replications. The vertical axis is the response time, or the time it takes to complete a single I/O over distance for the two types of replication. The horizontal axis is the relative I/O rate for a given separation distance. Only where the application I/O rate is below saturation (as shown in the shaded area) will asynchronous replication respond faster than synchronous replication.

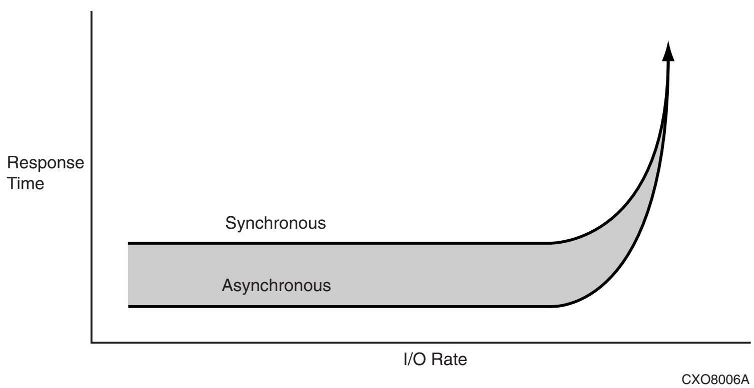

Figure 12: Asynchronous versus synchronous replication saturation

Figure 12 is both empirically and intuitively correct because Continuous Access EVA is replicating writes, and it takes resources (time and buffer space) to complete any write. Whether the replication write is performed before the write is acknowledged (synchronous), or after the write is acknowledged (asynchronous), both use the same buffers to move the data to the destination array.

Another way to consider this is that both types of I/O (synchronous and asynchronous) are queued at the local site. This queue is of a finite size and the peak rate at which it can be emptied is limited by the separation distance, not the type of replication.

Because both synchronous and asynchronous replication saturate at approximately the same rate, as shown by Figure 12, in disaster-tolerance planning environments the recommendation is always in favor of synchronous mode, for data protection, when all other factors such as peak I/O rate are equal. Not only does synchronous replication provide for data protection, asynchronous mode offers no additional throughput or performance capabilities over synchronous mode. In addition, both synchronous and asynchronous are supported at the maximum separation distance. This differs from other solutions where asynchronous is required when the separation distance is past a typical metropolitan threat separation.

Asynchronous CA EVA does not use the write history log for its buffer. Starting with VCS V3.02, a new asynchronous replication request is temporarily changed into a synchronous request whenever the replication buffer is full. Therefore, HP recommends the use of synchronous replication whenever possible.

Note: Because actual values depend on the size of the I/O and the intersite distance, only the relative relationship between response time and I/O rate is shown in Figure 12.

Bidirectional solution

Bidirectional replication allows for a storage system to contain both source and destination virtual disks, where these virtual disks belong to separate or unique DR groups. As shown in Figure 13 you can configure some data replication (DR) groups to replicate data from storage system A to storage system B, and other unrelated groups to replicate data from storage system B back to storage system A. This setup has no effect on normal operation or failover policy and has the advantage of allowing for the destination storage system to be actively used while also providing a disaster-tolerant copy of the other site’s data.

If the business needs require bidirectional data transfers, you must determine the effect on the intersite links.

For example, consider the bandwidth as two unidirectional flows, and then add the two flows together to get the worst-case bandwidth requirements in either direction. The worst-case scenario occurs during recovery after failover and should be used in delivering intersite bandwidth requirements.

Bidirectional replication also can be configured so that servers at the remote site, while performing secondary tasks, are ready to support the source applications if the local site is damaged. In addition, the remote site servers can be set up to handle some of the local site application load, if the application load is easily divided. Secondary tasks that can be performed by the remote site include backup, generating reports, and data mining.

Figure 13 demonstrates the bidirectional concept.

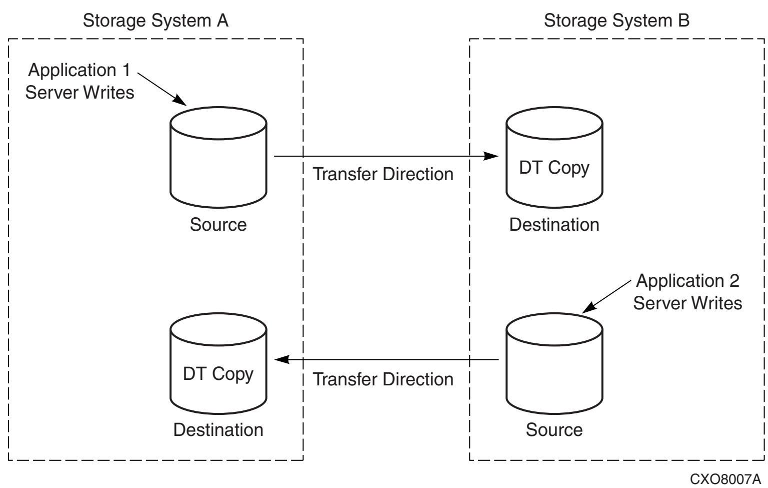

Figure 13: Bidirectional concept

Although multiple relationship configurations are supported, some results may not be practical (see Figure 14).

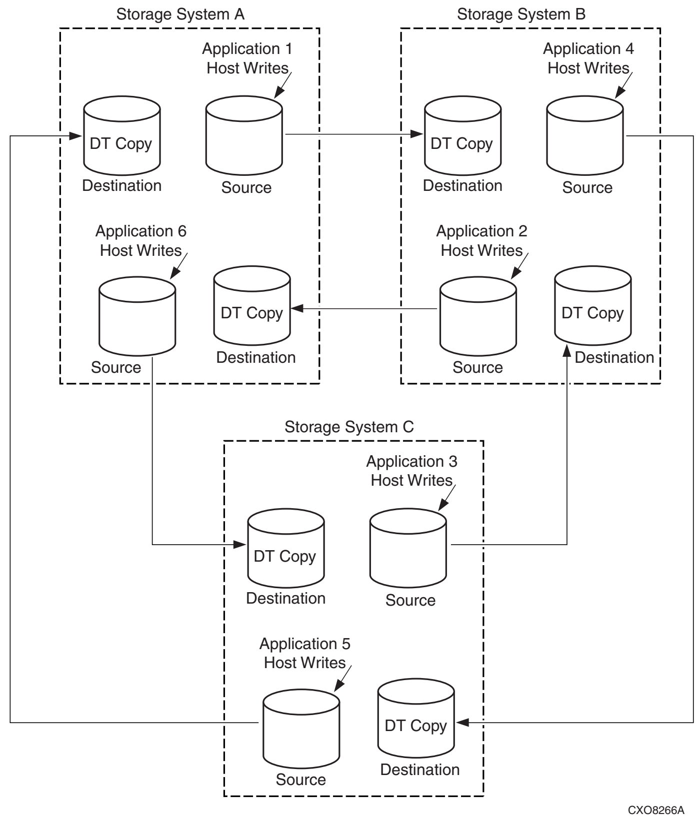

Figure 14: Multirelationship replication concept

Continuous Access EVA Design Considerations

3

This chapter describes the process for reviewing the design of the physical EVA storage that is part of the overall Continuous Access EVA solution. Considerations include the trade-off of availability and performance versus capacity.

Topics include:

Definitions

A *disk group* is the collection of physical disks or spindles that make up a single failure domain within the total capacity of the storage array. A single disk group can contain from 8 to 240 disks. An *array* can contain up to 16 disk groups, although this is limited by the total number of drives available, divided by 8. A *virtual disk* is carved out of the available space within a disk group and is striped across each member of the disk group, where possible. Virtual disks can be one of three HP VersaStor-enabled virtual RAID (Vraid) types:

- **Vraid0**—HP VersaStor-enabled virtual RAID 0, or stripe set (not recommended in Continuous Access EVA environments)
- **Vraid1**—HP VersaStor-enabled virtual RAID 1+0, or mirrored stripe set
- **Vraid5**—HP VersaStor-enabled virtual RAID 5, stripe set with parity

Copy sets are the virtual "joining" of two virtual disks, one on the local array and one on the remote array. A source and destination pair of *DR groups* contains from one to eight copy sets.

Disaster-tolerant virtual disks

As discussed in Chapter 1, a single Continuous Access EVA-based storage array with VCS V3.01 or higher supports a maximum of 128 copy sets, where a copy set is a logical storage unit whose contents are replicated to another storage unit on another array.

Although 128 copy sets are supported, there are reasons that justify not using all the available copy sets in an array. For example, an application can generate multiple I/O streams using the entire I/O bandwidth of the controller and only require five copy sets. Also, the size of a single virtual disk can require most of the available capacity within the storage array cabinet. Finally, some smaller Continuous Access EVA installations may not need all 128 copy sets. Each copy set can exist in any supported Vraid0, Vraid1, or Vraid5 format. Both the source and destination are of the same type.

Note: In a CA EVA configuration, HP does not recommend the use of Vraid 0 at any time and Vraid 5 on arrays with fewer than eight drive enclosures. See "Disk group and virtual disk availability considerations" on page 59 for additional information.

Non-disaster-tolerant virtual disks

A Continuous Access EVA-based storage array can also provide nonreplicated (non-disaster-tolerant) virtual disks. These sets of disks are not replicated to the remote site and can be used for almost any purpose. For example, each source virtual disk can be cloned as either a point-in-time snapshot or a Snapclone copy. The Snapclones are stored in the non-copy set virtual disks. Similarly, a Snapclone or snapshot of each destination virtual disk can be taken at the remote site.

Finally, the non-copy set virtual disk can be used for temporary or scratch space. Each non–copy set virtual disk can exist in any supported Vraid0, Vraid1, or Vraid5 format. Due to the effect on performance, HP does not recommend placing the page-swap file on a storage array that is also performing data replication.

Non-disaster-tolerant virtual disks also can be used for offsite backup. Backup of the replicated data at the remote site (called *remote vaulting*) is accomplished by first taking a snapshot of the data and then using a standby server to move the data from the snapshot to the backup media.

This process can occur at either site, but having the data onsite ensures faster restores by the storage array network back to the local site. Remote vaulting does not provide faster restores like onsite data does.

Up to seven active snapshot volumes per virtual disk, eight per DR group, are supported at any one time. The size of the snapshot depends upon the type—*demand-allocated* snapshots are efficient in their use of disk space, while *fully allocated* snapshots and Snapclones are inefficient. As the data changes in the original virtual disk, the snapshot or Snapclone volume grows or is updated to replicate the original volume.

Be aware that the process of creating a snapshot or Snapclone uses some of the same resources that are also used in the replication of that same virtual disk. Therefore, try to limit the number of Snapclone operations performed during times when peak replication performance is needed. With VCS V3.02, you can choose the Vraid type for both the Snapclone and the snapshot, as well as the location of the Snapclone.

Disk groups

This section explains how to design the storage side of a Continuous Access EVA solution. The first topic, "Disk group performance design," introduces the design parameters and discusses their impact. The second topic, "Disk group capacity design," offers a formula that determines the minimum number of disk drives needed to provide a known amount of usable disk space for a given level of redundancy or performance requirements. The third topic is on "Write history logs." Finally, the last topic covers "Fabric port count results," and discusses the factors that influence application performance from a storage perspective.

Caution: A disk group must contain only disks of the same size and speed, otherwise performance can be degraded. Additionaly, space is reserved based on the capacity of the largest drive within the group. Disk groups *must* contain only one type of drive (online or near-online).

Disk group performance design

Performance is directly related to the number of spindles that can be used to spread the data access across. The more spindles, the higher the aggregate data rate across the spindles. The converse is also true: the fewer the spindles, the lower the aggregate throughput across the group of spindles.

In Continuous Access EVA, this is not always true. Because of replication to a remote array, there is a point where adding more spindles does not increase the maximum write rate, only the maximum random read rate. If an application has a high percentage of random reads compared to writes, then a large number of spindles in the disk group is appropriate. If, however, there is a high percentage of writes compared to reads, replication limits the performance more than a limited number of spindles. Likewise, sequential access, either read or write, is limited by the per-drive performance more than the number of spindles in the disk group.

Another aspect of disk group design is the number of virtual disks that share the disk group. If only a few virtual disks share the disk group, and they all have small throughput requirements compared to a single drive, then the minimum number of drives needed to contain the total combined capacity of the virtual disks plus spare capacity is enough.

If the virtual disks that share a disk group have a large random read requirement compared to the number of writes, an optimal disk group should contain many more small-capacity drives rather than a few large-capacity drives, so that the aggregate I/O requirements of the virtual disks is less than the aggregate available from the disk drives.

Another way to view this problem is to consider that the average Fibre Channel disk is able to process approximately 125 to 175 requests per second, or on average, 150 requests per second. This variation in the actual performance is due to the size and locality of the requests, the number of servers issuing requests, and the rotational speed of the drive. Within the EVA, because a virtual disk is striped across all members of the disk group, the resulting Vraid0 bandwidth of a virtual disk is the product of the average single disk request rate times the number of drives in the disk group. The maximum Vraid1 write request rate is one half that of Vraid0, due to the extra write for the mirror. Similarly, the maximum Vraid5 write request rate is four-fifths (80%) that of the Vraid0 rate, due to the additional space needed by the parity write. The Vraid type does not impact the read rate as it does the write rate.

Therefore, given a request rate and a Vraid type, the minimum size of a disk group can be calculated. Consider, for example, an application that is expected to process an average of 1500 random write requests per second. Then, on average, the Vraid0 data should be striped across at least 1500 divided by 150, or 10 spindles. For Vraid1, the number of spindles doubles to 20, due to the 2 times the writes needed for the mirrors. For Vraid5, the number of spindles is 5 times 10, or 50, to retain the same effective request rate to the controller, and due to the two reads and two writes for every update of a Vraid5 data block.

HP recommends choosing the maximum number of drives based on comparing the number of drives needed for performance with the number of drives needed based on capacity.

Caution: Failure to provide a sufficient number of drives to satisfy performance expectations can result in performance degradation, including loss of data.

When planning the layout of the disks, round up a bit to fill a vertical column across all of the disk enclosures attached to a storage array. For example, if you are using an EVA rack with 12 enclosures, order disks and assign them to disk groups in blocks of 12. Similarly, if there are only four enclosures, then order and install in groups of four (also subject to the rule that there is a minimum of eight drives in a disk group). There must be an even number of drives in a disk group.

Disk group and virtual disk availability considerations

Before determining how many disk drives are needed for a particular solution, it is important to consider the availability of the three Vraid types. Note that each virtual disk, regardless of Vraid type, is built from many small data chunks. Each of these chunks of data is striped across an average of 8 physical drives, although the number can vary from 6 to 12, depending on the total number of drives in the disk group. How this is done is beyond the scope of this document. The point of this section is to understand how the number of drive enclosures affects the availability of a virtual disk for a given Vraid type. The more drive enclosures that are available to be used in building a disk group, the higher the reliability of the virtual disks in the disk group.

Least reliable are virtual disks built using Vraid0 technology, because there is no redundancy built into the data stripes and the failure of any one drive within such a disk group causes the loss of all Vraid0 disks within that disk group.

Caution: HP strongly recommends that you limit the use of Vraid0 virtual disks to nonreplicated temporary scratch disks, because the loss of a single disk within the group can result in the loss of data.

Virtual disks built up using Vraid1 technology can usually survive the loss of up to half the drives in a properly configured disk group. This can occur when one of the disk enclosures fails within a storage system. Therefore, for the highest availability of a virtual disk, HP recommends using Vraid1 virtual disks whenever possible.

The middle ground in availability are virtual disks built using Vraid5 technology, because they can survive the loss of some disks within the disk group. Within a properly configured disk group, Vraid5 disks can survive the loss of a disk enclosure if there are eight or more enclosures in the storage array. The eight-enclosure rule comes from a properly configured array.

For example, consider one of the largest EVA arrays available, the EVA5000 model 2C12D which contains two HSV110 controllers and 12 disk enclosures. Twelve enclosures usually provide built-in redundancy for Vraid1 and Vraid5 virtual disks to survive the loss of at least one drive enclosure. This is provided that the disks within each disk group are spread evenly across all 12 drive enclosures.

Consider now the smallest EVA5000 model—the 2C2D storage system. It contains two HSV110 controllers and only two disk enclosures. If one of the two enclosures fail, only the Vraid1 virtual disks are accessible. Due to the large number of failed drives, expect loss of access to data in Vraid0 and Vraid5 virtual disks. Therefore, in cases where the array contains fewer than eight drive enclosures, HP recommends using only Vraid1 virtual disks.

Best practice rules

When populating drives into an EVA cabinet there are four rules to consider:

- Rule One:
  - For performance and redundancy reasons, the number of drives of a given type (Enterprise Class or near-online and online) in a shelf enclosure should not differ by more than one from any other enclosure in the EVA subsystem.
  - Balance the number of similar drives across all shelves.
  - Remember that disk groups are made of drives of one type and are more robust if filled with the same number of drives per shelf.
  - Stack drives vertically in multiples of 8 completely filling vertical columns as shown in Figure 15 and listed in Table 18 on Page 61.

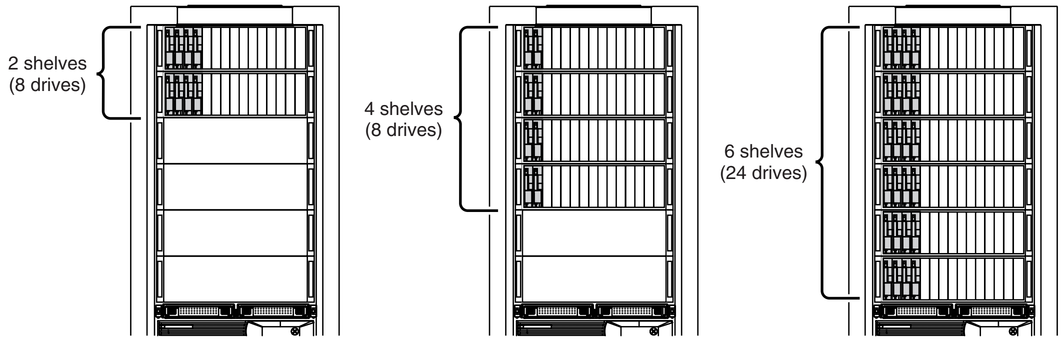

Figure 15: Drive placement best practices

Table 18: Drive requirements

Number of Shelves	Drive Multiples of	Minimum Drives Required for best availability
2	8	8
4	8	8
6	8	24
8	8	8
10	8	40
12	8	24
14	8	56
16	8	16
18	8	72

- Rule Two:
  - — Separate drives of different types, sizes, and speeds into their own group.
  - — The number of drives of a given size and speed in a shelf enclosure should not differ by more than one from any other enclosure in the EVA subsystem.
- Rule Three:
  - — When adding drives to, or removing drives from, an existing configuration, follow rules one and two for drive placement.
  - — Add drives of similar types to shelves that have empty bays.

- Rule Four:
  - When adding or removing a drive enclosure, follow rules one and two for drive placement after the changes are made.

Disk group capacity design

When determining your usable disk space and the quantity of drives needed to achieve a desired storage capacity, be aware that the full capacity of a drive is not available for data storage. For instance, a 36-GB drive yields a formatted capacity of 33.91 GB. In addition, a very small percentage (approximately 0.2 percent) of the disk group is used by the controller for storing disk group metadata. So in this example, a disk group with ten 36-GB drives actually equals:

(33.91 GB x 10 drives) x 99.8% = 338.4 GB

not the expected 360 GB. In practice, the actual overhead percentage varies according to the size of the disk group and the protection factor.

An overhead is also applied to the available storage space. For instance, when creating a virtual disk from the disk group in the previous example, the maximum Vraid0 capacity available is actually

338.4 GB x 99.6% = 337 GB

About 0.4 percent is space used for storing the virtual disk’s metadata (the actual overhead percentage increases significantly with the size of the disk group). Performing this same calculation for Vraid5 yields 270 GB, and for Vraid1 yields 169 GB of available space.

You can use the formula in Figure 16 to determine the minimum number of drives required to achieve a desired available storage space in a single disk group. Note that the Vraid redundancy type can vary within a disk group. The formula also accounts for the disk group protection level. The formula must be repeated for subsequent disk groups.

$$\mathrm{DC} = \frac{(\mathrm{usableVraid0} \times 538) + (\mathrm{usableVraid5} \times 673) + (\mathrm{usableVraid1} \times 1076)}{(\mathrm{DiskCapacity} \times 476)} + (\mathrm{PL} \times 2)$$

Figure 16: Disk drive formula

In Figure 16, the constants 538, 673, 1076, and 476 establish usage ratios that include reserving space for metadata and system overhead. DC is the Disk Count and PL is the Protect Level.

Example 1: The drive size is 72 GB and the protection level is 2. If you need 0.5 TB of Vraid0 space (usable Vraid0), 1 TB of Vraid5 space (usable Vraid5), and 1.5 TB of Vraid1 space (usable Vraid1), using 72 GB drives, with double protection level, the minimum number of disk drives (DC) would be:

$$\mathrm{DiskCount} = \frac{(500\mathrm{GB} \times 538) + (1000\mathrm{GB} \times 673) + (1500\mathrm{GB} \times 1076)}{(72\mathrm{GB} \times 476)} + (2 \times 2) = 79\mathrm{disks}$$

Always configure disk groups with an even number of drives. A system can never be configured with a disk group of fewer than 8 drives. Disk groups are most effective if the number of disks in the group is both an even multiple of the number of drive enclosures and 8. Therefore, in this case by rounding up to the next multiple of 8, the 79 becomes 80 drives.

Because more free space is better, when sizing disk groups consider that the efficiency of the EVA virtualization algorithms improves with the amount of free space available.

Note: Ten percent or more free space allows for the most efficient operation of the virtualization algorithms, including the even distribution of data across all members of the disk group.

Example 2: You need to know how many 36-GB disks to purchase for your EVA. You want three different virtual disks, with each virtual disk in its own disk group, all with a failure protection level of double. You want one virtual disk of each Vraid type and they must be 750 GB in size. In these calculations, DC equals the required number of drives, calculated using the formula in Figure 16.

Need 750 GB Vraid0:

$$\mathrm{DC} = \frac{750 \times 538}{36 \times 476} + 4 = 28\mathrm{drives}$$

Because 28 is not an even multiple of 8, the correct result requires 32 disk drives.

Need 750 GB of usable Vraid1:

Because 52 is not an even multiple of 8, the correct result requires 56 disk drives.

$$\mathrm{DC} = \frac{750 \times 1076}{36 \times 476} + 4 = 52\text{drives}$$

Need 750 GB of usable Vraid5:

$$\mathrm{DC} = \frac{750 \times 673}{36 \times 476} + 4 = 34\text{drives}$$

Because 34 is not an even multiple of 8, the correct result requires 40 disk drives.

Before rounding up, the total is 28 + 52 + 34 = 114 drives, which, when rounded to the nearest multiple of 8, is 120 drives. After rounding up each individual result, the total is 32 + 56 + 40 = 128. The extra drives allow for higher performance and better efficiency of the disk group, as described in the previous section on availability.

Note: The maximum capacity of an EVA is 32 TB (base 2), 35 TB (base 10), or 240 drives, whichever comes first.

Write history logs

Additional space is needed for the write history log that is used by a DR group whenever there is a problem with the intersite links. These logs can range in size, starting at 136 MB and growing up to 2 TB if enough disk space is available. The controller places the log in the disk group with the most free space. This can be the same disk group containing the DR group, or another one. In either case, the space must be included in the disk space planning. Logs are of type Vraid1.

Note: With VCS V3.02 or higher, the write history log is automatically placed into a disk group containing near-online disk drives, if one exists. Since this is done at the creation of the DR group it is necessary to recreate DR groups to move the log from an online disk group to a near-online disk group.

VCS V3.02 allows you to create disk groups using near-online disk drives in addition to online disk drives. Disk groups created using near-online disk drives are a more cost effective way to hold the write history logs for all DR groups that exist on an array. VCS V3.02 automatically selects a near-online-based disk group (should at least one exist at the time that the DR group is created) for placement of the write history log.

The following four cases illustrate the process that VCS uses to select a disk group for the write history log:

- **Case 1:** The array contains one defined disk group.

  Action: VCS automatically places the write history log into the defined disk group. The write history log will not move if additional disk groups are created after the DR group is created.

- **Case 2:** The array contains one near-online disk group and more than one online disk group.

  Action: VCS automatically places the write history log into the near-online disk group. The write history log will not move if additional disk groups are created after the DR group is created.

- **Case 3:** The array contains multiple near-online disk groups.

  Action: VCS picks the near-online disk group containing the most free space. If more than one near-online disk group has the same amount of free space, VCS alternates between the members, attempting to level out the use of the free space. The write history log will not move if additional disk groups are created after the DR group is created.

- **Case 4:** The array contains multiple online disk groups and no near-online disk groups.

  Action: VCS picks the online disk group containing the most free space. If more than one online disk group has the same amount of free space, VCS alternates between the members, attempting to level the use of free space. The write history log will not move if additional disk groups are created after the DR group is created.

Fabric port count results

The final trade-off to consider is the number of storage arrays and the number of servers used by the applications. In Continuous Access EVA V1.1, the maximum size of a SAN is limited to 28 switches per fabric for B-series switches, 16 switches per fabric for C-series switches, and 24 switches per fabric for M-series switches. There are two fabrics and a minimum of two sites, for a maximum of 56 B-series switches, 32 C-series switches, or 48 M-series switches, split between the two fabrics and at least two sites. These switches are distributed across the fabrics in one of several topologies, as described in chapter two of the *HP StorageWorks SAN Design Reference Guide*.

Fabric ports are consumed on each fabric at the rate of one port per Fibre Channel adapter (FCA) pair (or dual-port FCAs) and two ports per storage array. Refer to chapter 3 of the *HP StorageWorks SAN Design Reference Guide* for detailed fabric rules, such as the mixing of 1-GB and 2-GB switches, maximum number of ports, or mixing switch vendors.

The limits for server-to-storage array hops depend on the type of switches used (at the time of publication). Contact your local HP Representative for current limits.

- For B-series switches:
  - Switch count is limited to 28 per fabric, split between the two sites.
  - The source storage array should be located no farther than seven switch-to-switch hops from the server issuing the I/O.
  - The destination storage array should be located no farther than seven switch-to-switch hops from the array issuing the replication I/O.
- For C-series switches:
  - Switch count is limited to 16 per fabric.
  - The source storage array should be located no farther than seven switch-to-switch hops from the server issuing the I/O.
  - The destination storage array should be located no farther than seven switch-to-switch hops from the array issuing the replication I/O.
- For M-series switches:
  - Switch count is limited to 24 per fabric, split between the two sites.
  - The source storage array should be located no farther than three switch-to-switch hops from the server issuing the I/O.
  - The destination storage array should be located no farther than three switch-to-switch hops from the array issuing the I/O.

To determine the number of open ports required for the solution, add together the number of FCA and controller ports on each fabric at each site. Remember to account for all interswitch and intersite links when determining the number of open ports available for a given fabric topology.

For symmetric solutions, allow one hop for the intersite link and split the remaining hops evenly between the two sites.

Note: The *HP StorageWorks SAN Design Reference Guide* provides additional rules and limitations, such as maximum port count.

Continuous Access EVA Configurations

4

This chapter describes the supported basic Continuous Access EVA configurations, and the supported changes that you can make to them. The design rules for each of the subsequent configurations build on the basic Continuous Access EVA-over-fiber configuration.

Topics include:

Basic Continuous Access over fiber

Continuous Access EVA over fiber is the most basic of the Continuous Access EVA configurations; all others build upon it. As shown in Figure 17 on page 71 the cabling in this configuration supports two redundant fabrics, where the first FCA in server A is connected to switch W, and the second FCA in server A is connected to switch X. The top controller of the storage array on the left is attached to switch W and switch X, and the bottom controller is also attached to switch W and switch X.

At the right side of the figure, the backup server and the storage array are wired the same way as at the left-hand site. Using the same switch ports for the same functions at both sites reduces confusion during a disaster or debugging. HP also recommends naming the two fabrics to distinguish them. For example, name them TOP and BOTTOM, BLACK and GRAY, and so on.

This dual-fabric SAN provides no single point of failure (NSPOF) at the fabric level. For example, broken cables, switch updates, or an error in switch zoning can cause one fabric to fail, leaving the other to temporarily carry the entire workload. At the time of publication, up to 28 B-series switches, 16 C-series switches, or 24 M-series switches are supported per fabric. For the most current limits, contact your local HP Representative. Non-Continuous Access EVA servers and storage are allowed on each fabric, if each is kept in a zone separate from the Continuous Access EVA solution space.

Total solutions larger than what is possible with a single supported solution are also supported. Each of the smaller solution instances must exist within a single management zone that conforms to all the requirements outlined in the section "Basic Continuous Access EVA configuration rules" on page 73. The combination of two or more solution instances must not exceed the maximum configuration in the *HP StorageWorks SAN Design Reference Guide*.

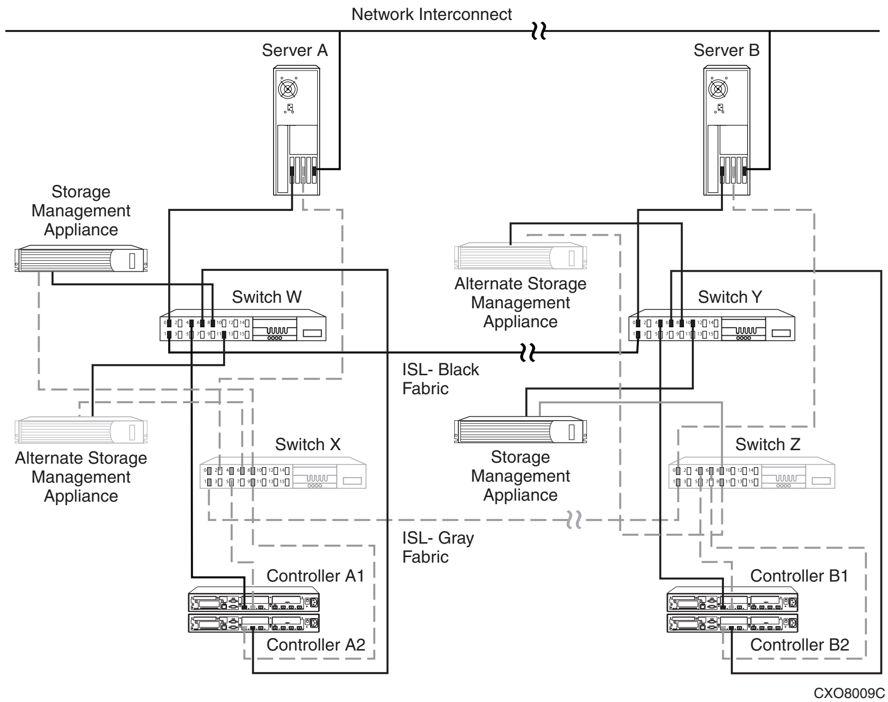

Figure 17: Basic Continuous Access-over-fiber configuration

Cabling

Figure 18 on page 72 shows supported cabling. The basic rule is that the first or left-hand port on the top controller is cabled to the first fabric, and the other port of the same controller is cabled to the other fabric. The other (bottom) controller is cabled so that the left-hand port is attached to the second fabric, while the second port is cabled to the first fabric; the opposite of the first (top) controller. Even though it does not matter which switch ports are used, symmetry is recommended. If there is a fabric failure, the LUN stays on the same controller but moves to the other fabric. Any other controller-to-fabric cabling scheme is not supported.

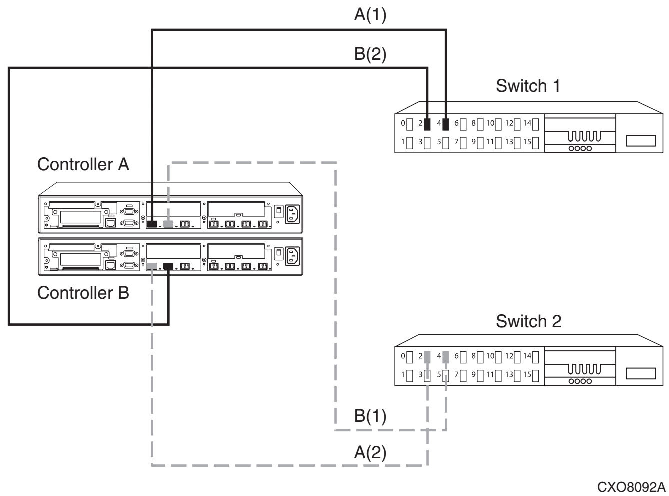

Figure 18: Supported cabling

The cabling in Figure 19 is not supported because both ports of a controller are on the same fabric, and the LUN changes controllers if a fabric failure occurs.

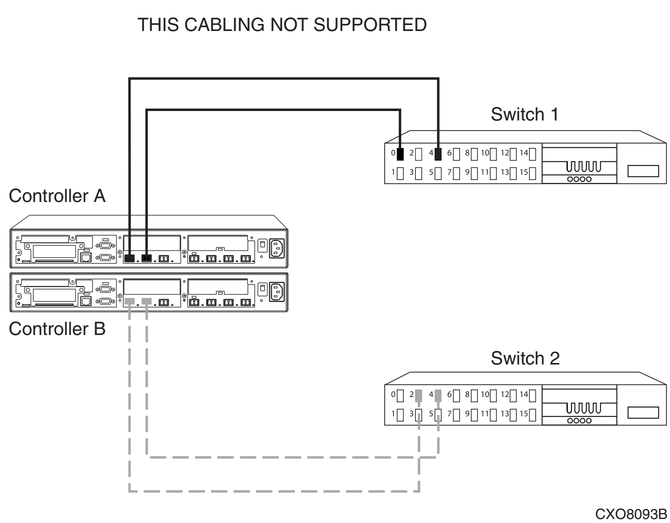

Figure 19: Example 1 of cabling that is not supported

The cabling in Figure 20 is not supported because both port 1s and port 2s share the same fabric. This inhibits static load balancing by DR group and limits failover operations.

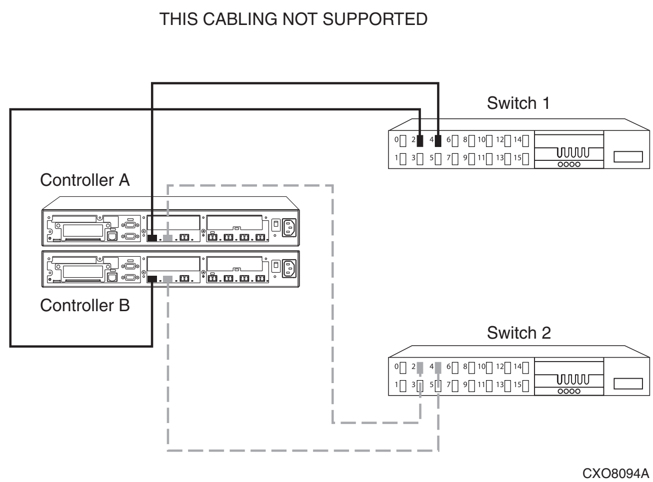

Figure 20: Example 2 of cabling that is not supported

Basic Continuous Access EVA configuration rules

Consider the following requirements when designing your Continuous Access EVA system:

- At least two, but no more than 16, Continuous Access EVA storage systems can be split between the local and remote site. Each storage system dedicated to support a Continuous Access EVA configuration must have dual HSV controllers.
- The operating system on each server must either implement multipath support (as do OpenVMS and Tru64 UNIX) or support it by means of HP StorageWorks Secure Path software.
- The minimum HP StorageWorks Enterprise Virtual Array storage configuration supports up to 28 drives per storage array, with larger configurations supporting up to 240 drives (with expansion cabinet). Destination LUNs must have the same geometry and must be the same capacity as the source LUNs of the copy set.

- Source and destination disk groups do not require the same geometry (speed online and near-online) and need not be of the same capacity. However, the disk group with the fewer drives has a lower peak performance than the one with more drives. Disk groups supporting bidirectional replication should be symmetric in both size and performance of disks for consistent performance from either array.

Note: A disk group should contain only one model of physical disk.

- Both controllers within an EVA must have the same version (either version 3.00, 3.01 or 3.02) of the VCS Kit installed and running.

Note: Both EVAs must be running the same version of VCS.

- A minimum of 2 FCAs (or 1 dual-port FCA) are required per server to ensure that no single point of failure exists between the server and the storage system. And, on average, no more than 4 single-port, 2 single-port and 1 dual-port, or 2 dual-port FCAs per server.
- A maximum of 256 FCAs per storage system are allowed. The ports can be a mix of single- and dual-port FCAs. At two FCA ports per server, the 256 limit equates to a maximum of 128 servers.
- To maintain write order across the members of a DR group and to maintain a fail one/fail all model, all members of the DR group must be preferred to the same EVA controller and use the same FCA port pair. Therefore, on a system with multiple pairs of FCAs, all the virtual disks belonging to the same DR group are restricted to using only one FCA port per server.
- Each site must have one Continuous Access EVA documentation set and one storage array platform kit (appropriate for the array type and server operating system) per implemented operating system platform.
- One GBIC or SFP that is appropriate for the type of fiber optic cable being used is required per switch port connection.
- Each site must have at least one Storage Management Appliance; however, two are recommended for high availability of the management function.
- The designed configuration for switches must be supported by the *HP StorageWorks SAN Design Reference Guide*.

A third-party vendor is used to acquire and install all SMF optic cables, any MMF optic cables longer than 50 m, and WDM interfaces.

Maximum Continuous Access EVA-over-fiber configuration

As a maximum configuration, Continuous Access EVA over fiber supports up to 16 arrays split between the two sites, depending on the SAN topology. High performance SANs can be built using a *skinny tree* topology as defined in the *HP StorageWorks SAN Design Reference Guide*. Using a high-performance SAN, however, can reduce the maximum number of servers and storage arrays, due to the reduction in open switch ports.

A large port-count SAN can be constructed using multiple instances of smaller configurations, each in a separate management zone, subject to the limits in fabric sizes as defined in the *HP StorageWorks SAN Design Reference Guide*.

Note: See Table 17 on page 47 for limits based on separation distance.

Continuous Access EVA over fiber with long-distance GBICs and SFPs

Basic Continuous Access EVA over multimode fiber supports distances of up to 500 m at 1 Gbps, and up to 300 m at 2 Gbps.

Longer distances require the use of long-distance and very long-distance GBICs and SFPs. Long-distance GBICs and SFPs using single-mode 9-µm fiber can span distances up to 10 km. At the time of this publication, very long-distance GBICs and long-distance SPFs running on 9-µm fiber can support distances up to 100 km at 1 Gbps, or up to 35 km at 2 Gbps.

B-series switches may require the optional Extended Fabric License for this configuration.

Continuous Access EVA over WDM

As an option, Continuous Access EVA over fiber also supports the use of wavelength division multiplexing (WDM) instead of the long-wave or very long-distance GBICs.

Figure 21 shows a Continuous Access EVA-over-WDM configuration. At the time of this publication, HP supports Continuous Access EVA over any vendor's dense wavelength division multiplexing (DWDM) or coarse wavelength division multiplexing (CWDM) system, if the installation conforms to vendor specifications.

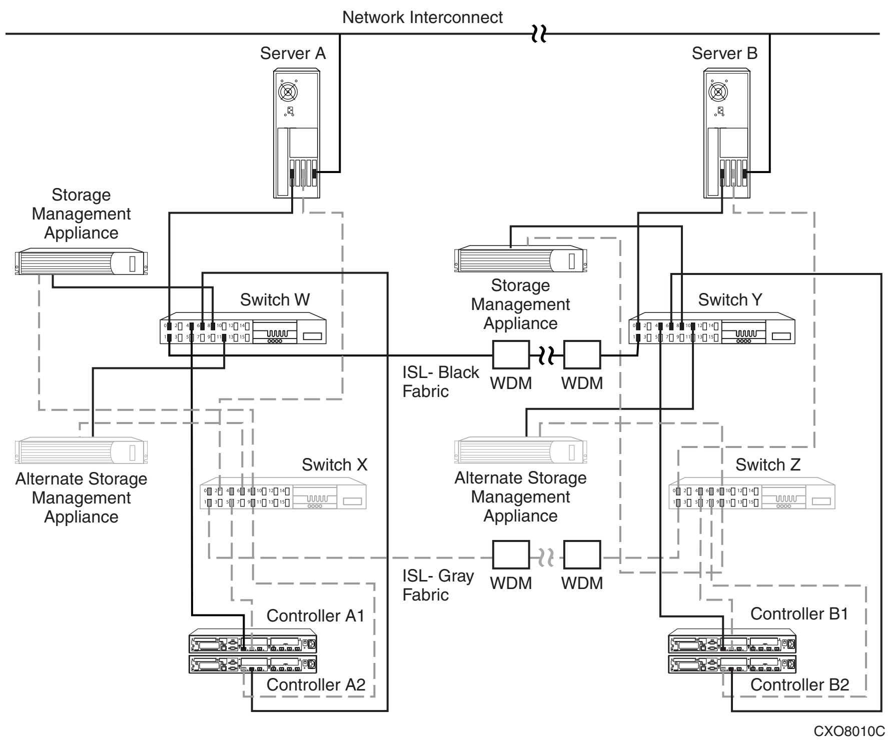

Figure 21: Continuous Access EVA-over-WDM configuration

The difference between the use of WDM and the basic solution is the replacement of at least one, if not both, long-distance GBICs and single-mode fiber with a multiplex unit, shortwave GBICs, and multimode fiber. For more information on Continuous Access over WDM, refer to the *HP StorageWorks Continuous Access and Data Replication Manager SAN Extensions Reference Guide*.

Additional configuration rules for Continuous Access EVA over WDM

Keep the following requirements in mind when you design your Continuous Access EVA-over-WDM system:

- Typically, one run per wavelength of multimode fiber is required to connect the switch to the WDM unit.
- If you are using older B-series switches, an Extended Fabric License may be recommended.

Extended Continuous Access EVA-over-IP configuration long-distance solution

The extended Continuous Access EVA-over-IP configuration is similar to the simple Continuous Access EVA configuration except for the use of Fibre Channel-to-IP gateways. Due to the dual fabrics, two gateways are required at each site—one per fabric, for a total of four per solution, dedicated to that solution.

Continuous Access EVA over IP has the same maximum configuration limits as those described in the section "Basic Continuous Access EVA configuration rules" on page 73. Multiple instances can share the same fabric as long as all components are in unique management zones and the ISLs are sufficient for all traffic flowing between the sites in a worst-case scenario.

Note: See Table 17 on page 47 for limits based on seperation distance.

Current versions of the Fibre Channel-to-IP gateways support direct connection to either 10/100-Mbps copper or 1-Gbps optical Ethernet. The Fibre Channel-to-IP (FCIP) gateway uses the intersite network bandwidth that is set aside for the storage interconnect. The IP tunnel that is created to support the FCIP traffic also provides enhanced security of the data because of the nature of IP tunnels. HP recommends designing the IP tunnels with enough bandwidth to carry all the intersite data traffic in case either link fails.

Packet loss within the IP network is tolerated and even expected, due to oversubscription or inadvertent congestion. As a result of these errors, a packet can be dropped without the sender being notified that the packet is now lost.

For every packet that is lost in the intersite network, that same packet must be retransmitted by the sender until received at the receiver or until a link timeout occurs. The effect of these retransmissions is seen as increased delay on the particular packet and a resulting decrease in the available bandwidth because it is not available for new packets. The greater the percentage of packets lost in the transfer, the lower the effective bandwidth and the longer a particular transfer will take. For example, using a maximum bit error rate (BER) of 1 in 10^{10}, one in approximately 500,000 2-KB data frames will require retransmission. At the other extreme, a 1 percent packet loss translates to losing 1.3 data frames in 100 2-KB packets, or some part of almost every 64-KB write due to network errors.

By way of comparison, Fibre Channel networks are designed for a BER of 1 in 10^{12}, or one in approximately 50,000,000 2-KB data frames. Because of this, HP recommends that both a maximum delay and maximum BER be specified in the network service contract.

Some wide area networks can be built using a ring architecture where one way around the ring is significantly shorter in time than the other (longer) way around the ring. Other wide area networks can also have two paths of different lengths, a shorter and a longer one. In either case, Continuous Access EVA supports these occasional drastic changes in the intersite delay, if the longest delay does not exceed an average of 100 ms. To accomplish this, the EVA storage system firmware periodically tests the intersite delay and adjusts the heartbeat rates, message time-outs, and outstanding I/O counts for optimum performance of an intersite link, based on the current intersite delay.

Figure 22 shows a Continuous Access EVA-over-IP configuration.

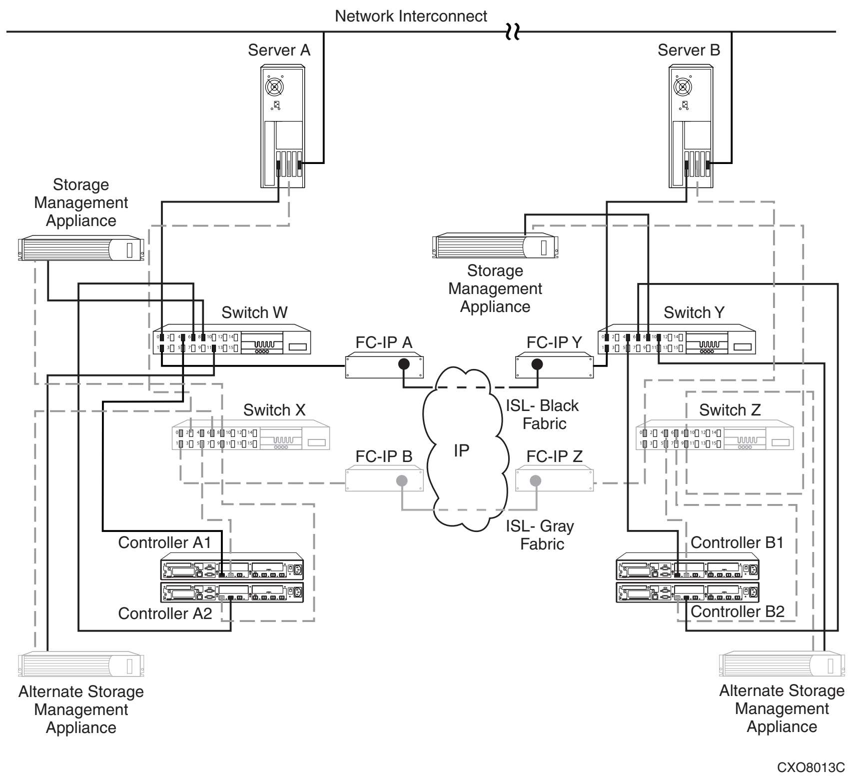

Figure 22: Continuous Access EVA-over-IP configuration

Additional configuration rules for Continuous Access EVA over IP

Consider the following requirements when designing your Continuous Access EVA system:

- Typically, one multimode fiber is required to connect the switch to the FCIP gateway.
- Some FCIP gateways are supported only on the older B-series switches and require the Remote Switch Key (vendor-dependent).

A third-party vendor is used to acquire and install all SMF optic cables, any MMF optic cables longer than 50 m, and the FCIP interface boxes.

Some IP gateways provide a mechanism to notify the fabric that connectivity to the remote gateway has been lost. Other gateways require the use of a fabric-based heartbeat to detect loss of the intersite IP network connection. Vendors that require the fabric heartbeat require installation of the Remote Switch Key license onto those two switches that directly connect to the IP gateway.

Refer to the *HP StorageWorks Continuous Access and Data Replication Manager Extensions Reference Guide* for more information.

Note: The Remote Switch Key is available only on the B-series switches. For those gateways requiring the Remote Switch Key, and on those switches where the Remote Switch Key is installed, do not enable suppression of F-Class frames. Doing so limits the supported size of the Continuous Access EVA-over-IP SAN to one switch per fabric at each site.

Note: No matter what type of site-to-site transport you use (IP, ATM, SONET, and so on), the FCIP gateway requires either a 10/100-Mbps copper or a 1-GbE optical interface into the local Ethernet network. The conversion from the local Ethernet to the long-distance network is expected to be performed by a customer-provided network router or gateway.

Continuous Access EVA over SONET

The extended Continuous Access EVA over ATM or SONET configuration is similar to the Continuous Access EVA over IP configuration except for the use of Fibre Channel-to-IP gateways. In Figure 22 on page 79 replace references to Fibre Channel-to-IP (FC-IP) gateways with SONET gateways.

Continuous Access EVA over ATM

Because no FC to ATM gateways are available at the time of publication, another approach is needed if a solution requires ATM as the transport between the two sites. That approach uses the same FC to IP gateways mentioned earlier, but the IP router is required to provide a blade that interfaces with the ATM network instead of the IP network.

Failover frequency

The planned or unplanned failover of one or more DR groups should not be performed more frequently than once every 15 minutes. The planned or unplanned failover of a controller should also not be performed more frequently than once every 15 minutes.

Continuous Access EVA stretched cluster support

Continuous Access EVA supports stretched Microsoft Cluster Servers (MSCS) running Windows 2000, Windows 2003, or Windows NT. In this configuration, shown in Figure 23 on page 82 half the cluster is at the local site and the other half is at the remote site. If the source server fails, MSCS fails over the application to the surviving server at the remote site and resumes operations using the local site storage.

Applications running in a stretched cluster in server failover mode incur a performance penalty because of the time it takes to read or write data across the intersite link. This performance penalty is directly proportional to the distance between the two sites. During testing, almost no additional effect was observed with separation distances up to 100 km. For more information on stretched cluster support, see the HP ProLiant HA/F500 website at:

http://h18000.www1.hp.com/solutions/enterprise/highavailability/microsoft/haf500/description-eva.html.

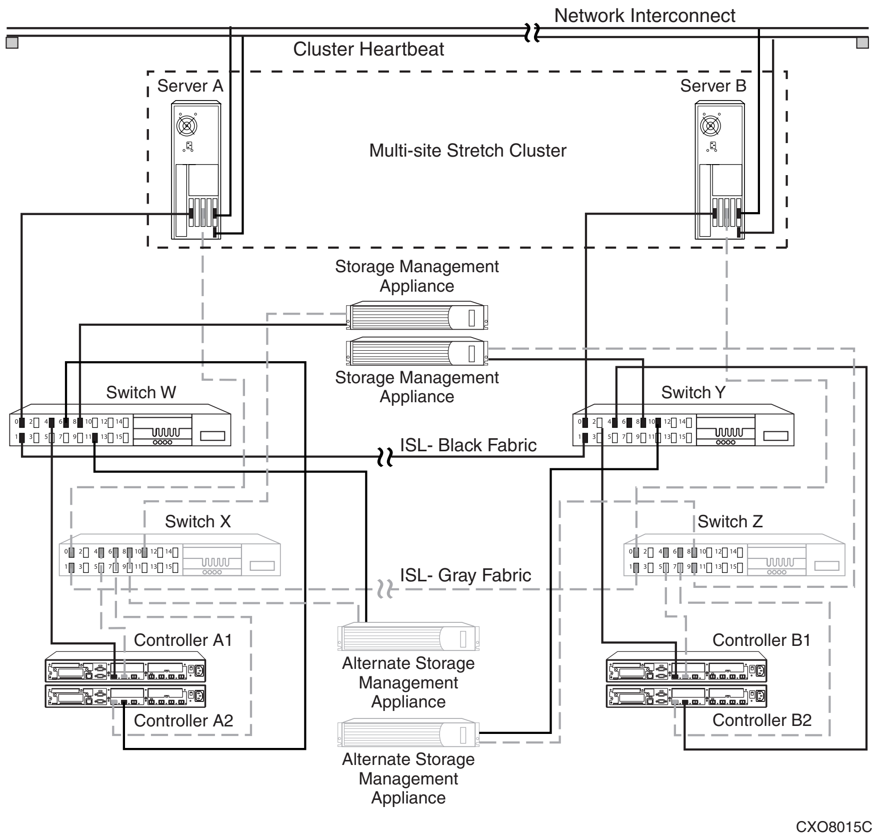

Figure 23: Continuous Access EVA stretched cluster configuration

Alternate configurations

The following configurations are supported but do not offer the same level of disaster tolerance and/or high availability as the "Basic Continuous Access EVA configuration rules" on page 73.

Single-fabric configuration

The single-fabric Continuous Access EVA solution is designed for small, entry-level tests or proof-of-concept demonstrations where some distance is needed between each of the two switches in the solution. This solution can also be used for producing copies of data needed for data migration or data mining.

Fabric zoning can be used to create two logical fabrics out of the one physical fabric. Fabric zoning is required to isolate servers as documented in the *HP StorageWorks SAN Design Reference Guide*. These two switches share one intersite link, leaving the remaining ports for servers, storage controllers, and a Storage Management Appliance. For example, if a 16-port switch is being used, the remaining 15 ports support up to:

- Four servers, one array, and one Storage Management Appliance.
- Two servers, two arrays, and one Storage Management Appliance.

An example of the single-fabric configuration using 16-port switches is shown in Figure 24.

Each of the switches shown in Figure 24 on page 84 can be replaced with any supported fabric topology as defined in the *HP StorageWorks SAN Design Reference Guide*. The same limits apply; that is, up to 28 B-series, 16 C-series, or 24 M-series switches are allowed in the single fabric.

All intersite links supported in the basic Continuous Access EVA are also supported in the single-fabric configuration. This means that the ISL can be direct fiber, a single WDM wavelength, or a single Fibre Channel over IP link.

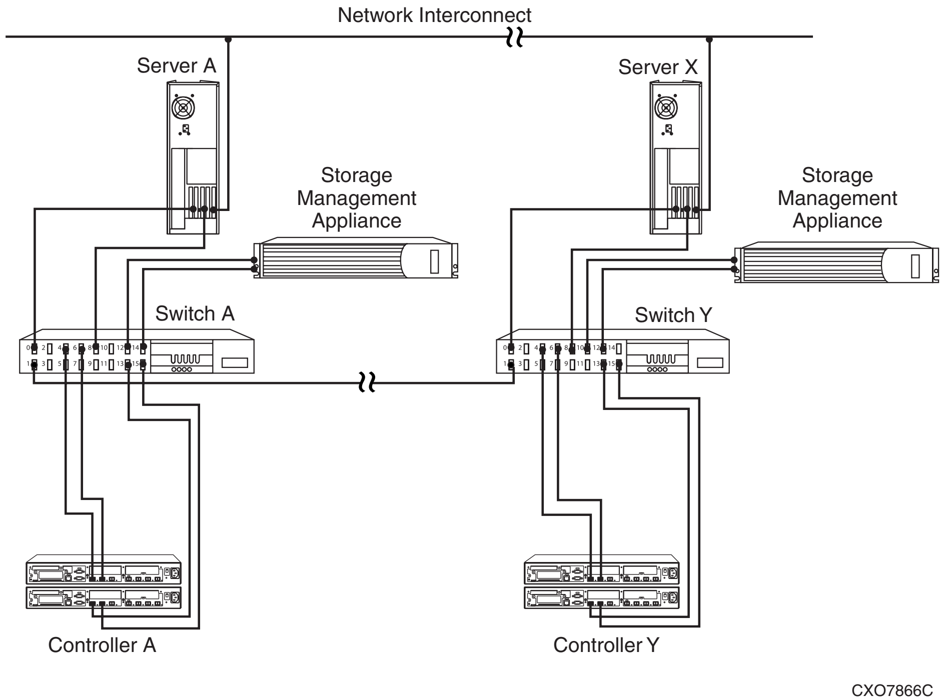

Figure 24: Single-fabric configuration

Single-switch configuration

The single-switch Continuous Access EVA solution is designed for small, single-site, entry-level tests or proof-of-concept demonstrations. This non-disaster-tolerant solution can also be used for producing copies of data needed for data migration or data mining. Dual FCAs and multipathing software are required. A 16-port switch can support a maximum of three servers, two storage arrays, and one Storage Management Appliance. Large switches support more servers and/or storage arrays if all FCA and array ports are connected to the same switch. Fabric zoning can be used to create the two logical fabrics used by Continuous Access EVA. Fabric zoning is required to isolate servers as defined in the *HP StorageWorks SAN Design Reference Guide*. An example of the single-switch solution is shown in Figure 25 on page 85.

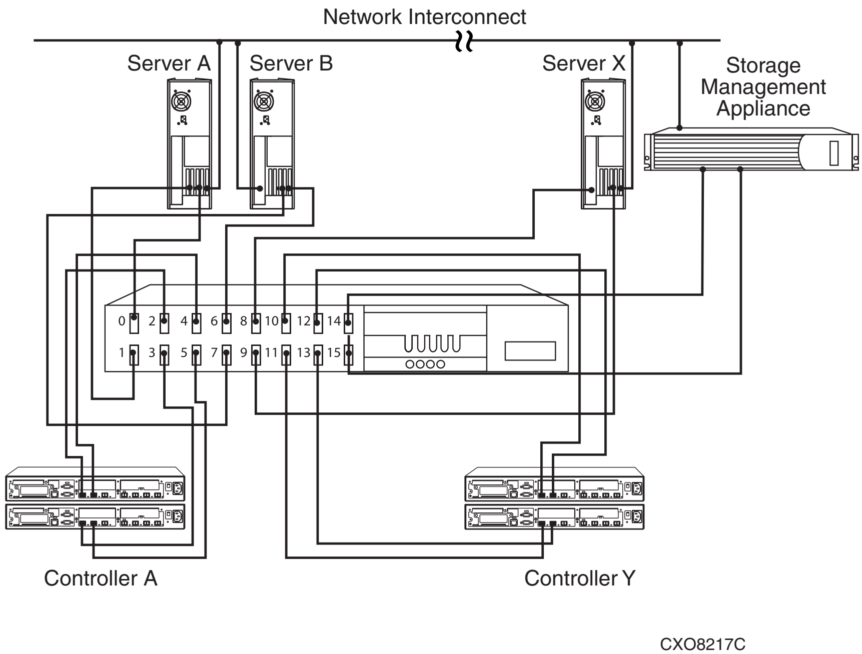

Figure 25: Single-switch configuration

In Figure 25, servers A and B are of one supported operating system and are clustered together using a supported cluster technology for that operating system. In this example, Server X is a single server running the same OS as the clustered servers A and B, and therefore is available as a backup to the cluster. As another example, Server X with a different OS can be a standalone server used for training on storage failover.

The single switch shown in Figure 25 can be replaced with any supported fabric topology as defined in the *HP StorageWorks SAN Design Reference Guide*. The same limits apply; that is, up to 28 B-series, 16 C-series, or 24 M-series switches are allowed in a single fabric.

Single FCA solution

A server containing a single FCA can be attached to any of the previously described configurations:

- Basic Continuous Access EVA, and its optional links
- Single fabric
- Single switch

This option allows the use of servers that only support one FCA due to slot restrictions, at the expense of reduced availability due to the single point of failure. To decrease repair time, HP recommends that you deploy some number of standby servers, each with a single FCA. If supported, each of these active and standby servers should be configured to boot from the SAN, so that any standby server could quickly replace any active server.

Note: Secure Path is required to mask the redundant path except when using HP Tru64 UNIX or HP OpenVMS.

Advanced configurations

Multiple replication relationships

Continuous Access EVA supports three types of multiple replication relationships:

- Fan out replication
- Fan in replication
- Cascaded replication

These relationships are between the individual arrays, where each relationship supports one or more DR groups not involved in another relationship. In other words, any one DR group and the copy sets within it can only belong to one relationship.

Fan out relationships: A —> B, A —> C

Figure 26 on page 87 shows an example of a fan-out relationship, where one DR group is being replicated from Array A to Array B, and another from Array A to Array C.

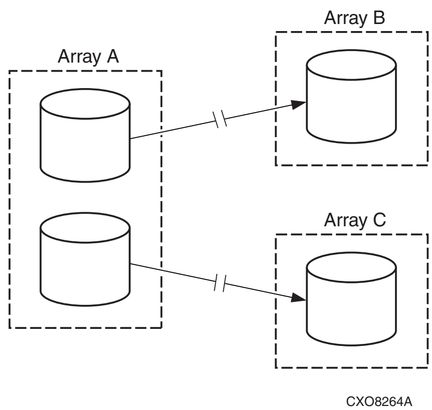

Figure 26: Array fan out relationships

Fan in relationships: A —> C, B —> C

Figure 27 shows an example of a fan-in relationship, where a DR group is being replicated from Array A to Array C and another from Array B to Array C.

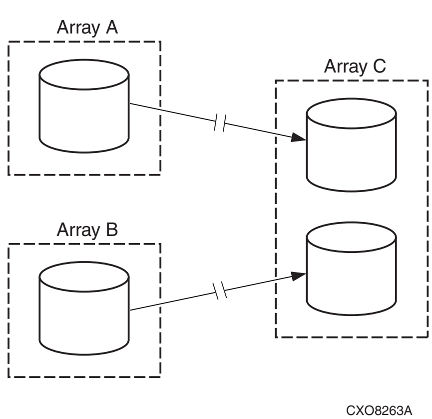

Figure 27: Array fan in relationships

Cascaded relationships: A—> B —> C

Figure 28 on page 88 shows an example of a cascaded replication, where one DR group is being replicated from Array A to Array B and another from Array B to Array C. In this case, the source disk for the Array B to C replication is a point-in-time copy of the destination disk in the Array A to Array B replication.

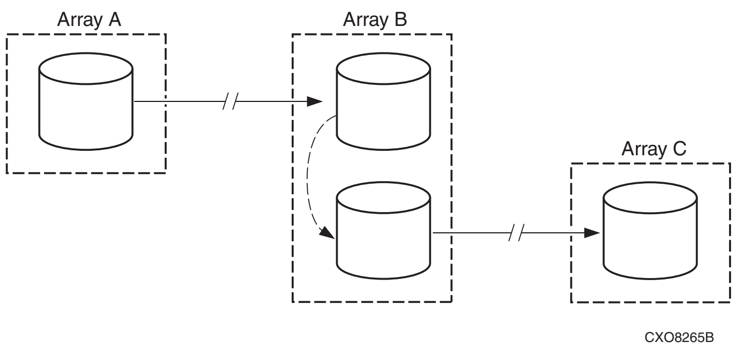

Figure 28: Cascaded array relationships

Bidirectional ring relationships

Figure 29 on page 89 is impractical in all but very specialized environments. In this case, there are three DR groups replicating clockwise, and another three replicating counter clockwise. None of the six DR groups are related, other than the source of one may have been a Snapclone of a destination of another on the same array.

Note: In multi-member DR groups, you may need to quiesce the application before starting the Snapclone operation to ensure data consistency across the members.

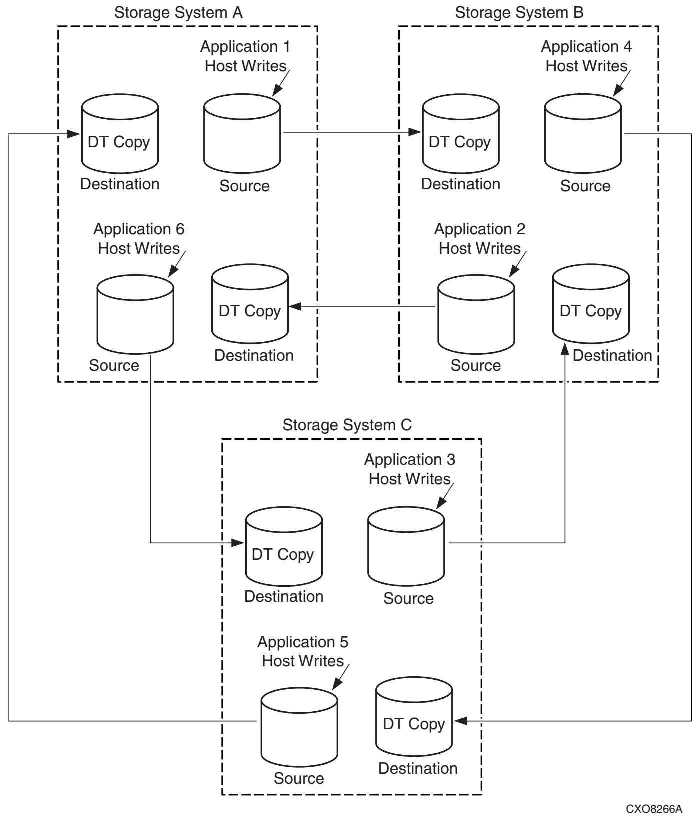

Figure 29: Bidirectional ring relationships

Operating System and Application Design Considerations

5

This chapter describes the process for reviewing your current application environments to determine if they will function in the Continuous Access EVA solution. It also discusses the capabilities that are available in operating systems and applications in a Continuous Access EVA environment. The same capabilities are not always available in non-Continuous Access EVA environments.

Topics include:

General operating system design considerations

Two concepts are important to understand in considering an operating system's design: booting from the SAN and bootless failover. The following sections explain both concepts.

Booting from the SAN

An important design consideration is whether to boot servers from the SAN and, if so, whether to replicate those boot disks using Continuous Access EVA. Currently, only HP OpenVMS, HP-UX 11i, HP Tru64 UNIX, and all supported Windows operating systems support booting from the EVA-based disks. With Continuous Access EVA, it is possible to replicate those boot disks to the remote EVA for use in recovery of the server and its applications, with the data. The only restriction is that you must place high-performance files, such as page and swap files, on storage within the server, rather than on the actual EVA-based boot disk. Otherwise, there is a potentially severe performance impact to the server caused by the replication of the writes to these files.

Note: If you replicate a boot disk with a defined IP address, HP recommends that both sites be in the same IP subnet so that you do not need to change the address after failing over to the backup copy of the system disk. Otherwise, you must change IP addresses after a server failover, such as during the network startup.

Bootless failover

Bootless failover allows destination servers to find the new source (after failover of the storage) without rebooting the server. This capability also includes the fail back to the original source without rebooting.

Operating system specific design considerations

Table 19 lists the hardware and operating systems supported by Continuous Access EVA at the time of publication. For the most current list, contact your local HP Representative.

Table 19: Continuous Access EVA-supported operating system versions and cluster support

Operating System	Version	Cluster Support
HP HP-UX	■ V11.0 and 11i v1 ■ HP StorageWorks Secure Path for HP-UX V3.0A with SP 1 or V3.0B with SP 1	■ HP MC/Serviceguard Clusters ■ VA.11.14 (4 node max.) ■ V11.0—SGV 11.14 ■ 11i v1—SGV 11.14 or 11.15
	■ 11i v2 ■ HP StorageWorks Secure Path for HP-UX V3.0B with SP 1 or V3.0C	■ HP MC/Serviceguard Clusters ■ VA.11.15 (4 node max.)
	■ V11.23 ■ HP StorageWorks Secure Path for HP-UX V3.0C or V3.0D	■ V11.14 ■ 11.11 ■ 11.23
HP OpenVMS	V7.2-2, V7.3-1, and V7.3-2	OpenVMS Clusters (96 node max.)
HP Tru64 UNIX	V5.1, V5.1a, and V5.1b	TruCluster (8 node max.)
IBM AIX	■ V4.3.3 ■ HP StorageWorks Secure Path for AIX V2.0C	None
	■ V5.1 and V5.2 ■ HP StorageWorks Secure Path for AIX V2.0D	■ AIX V5.1—HACMP V4.4 ■ AIX V5.2—HACMP V4.5 ■ 2 node max.

Table 19: Continuous Access EVA-supported operating system versions and cluster support

Operating System	Version	Cluster Support
Microsoft Windows NT Server	■ V4.0A, Service Pack 6a ■ HP StorageWorks Secure Path for Windows V4.0 required	MSCS V1.1 clusters (2 node max.)
Microsoft Windows 2000	■ V5.0, Service Pack 2, 3, or 4 ■ HP StorageWorks Secure Path for Windows V4.0B or V4.0C	■ MSCS V1.1 clusters (2 node max. ■ Advanced Server only
Microsoft Windows 2003 (32- and 64-bit)	■ V6.0 ■ HP StorageWorks Secure Path for Windows 4.0B required	■ MSCS V1.1 clusters (2 node max.) ■ 4 node max. on 32-bit ■ 8 node max. on 64-bit
Novell NetWare	■ V5.1, V6.0, and V6.5 ■ HP StorageWorks Secure Path for NetWare V3.0C	Novell Cluster Services (6 node max.)
Red Hat Linux (32-bit)	■ Advanced Server V2.1, QU3 2.4.9-e38smp and enterprise ■ HP StorageWorks Secure Path for Linux V3.0C	LifeKeeper Clusters (2 node max.)
	■ Advanced Server V3.0, QU2 2.4.21-15.ELsmp ■ HP StorageWorks Secure Path for Linux V3.0C, Service Pack 2	
Red Hat Linux (64-bit)	■ Advanced Server V2.1 QU3, 2.4.18-e43smp ■ HP StorageWorks Secure Path for Linux V3.0C	N/A
	■ Advanced Server V3.0, QU2 2.4.21-15.EL ■ HP StorageWorks Secure Path for Linux V3.0C, Service Pack 2	
Sun Solaris	■ V2.6, V7, and V8 ■ HP StorageWorks Secure Path for Sun V3.0A SP1, V3.0B, or V3.0C	■ Sun Clusters V2.2 ■ Veritas Cluster Managers V3.0 and V3.5 ■ 8 node max.
	■ V9 ■ HP StorageWorks Secure Path for Sun V3.0B SP1or V3.0C	

Table 19: Continuous Access EVA-supported operating system versions and cluster support

Operating System	Version	Cluster Support
SuSE Linux (32-bit)	■ SLES 8, 2.4.21-198smp, 2.4.21-215smp ■ HP StorageWorks Secure Path for Linux V3.0C	LifeKeeper Cluster (2 node max.)
	■ United Linux V1.0, 2.4.21-198smp, 2.4.21-215smp ■ HP StorageWorks Secure Path for Linux V3.0C	
SuSE Linux (64-bit)	■ SLES 8, 2.4.21-itanium2-smp ■ HP StorageWorks Secure Path for Linux V3.0C	N/A
	■ United Linux V1.0, 2.4.21-itanium2-smp ■ HP StorageWorks Secure Path for Linux V3.0C	

Note: For any operating system you use, refer to the OS-specific documentation to ensure that you use compatible versions of Secure Path drivers and HBAs.

HP

HP-UX

HP-UX supports booting from the SAN without a cluster. Bootless failover is not supported. HP UX supported file systems are VX and Veritas 3.5.

HP OpenVMS

HP OpenVMS supports booting from the SAN with and without clusters. With OpenVMS version 7.2-2, paths to the storage are not dynamic and are lost if there is a failure or other lost connection to the controller.

HP Tru64 UNIX

HP Tru64 UNIX supports booting from the SAN with and without clusters. In a failover situation, LUNs can be acquired but you may need to reboot the host if the LUNs can not be seen. HP Tru64 UNIX supports the Advanced file system and the UNIX file system.

IBM AIX

IBM AIX does not allow booting from the SAN, and does not support bootless failover. The supported file system is Journal file system (JFS).

Microsoft Windows

All Windows operating systems

Booting from the SAN is supported on all Microsoft operating systems. Bootless failover is not supported. The file system for all Microsoft operating systems is NTFS. HP recommends that you apply all Windows patches for security reasons.

Windows NT

When using Windows NT 4.0 to access an EVA running VCS V3.02, Qlogic drivers are not supported and Emulex cards support 4-4.82a16 drivers only.

Novell Netware

Novell Netware does not support booting from the SAN or bootless failover. Netware supports traditional Netware volumes and Novell Storage Services (NSS) file systems.

NetWare 6

If you use NSS logical volumes in a DRM configuration, note that Novell's Distributed File Services (DFS) allows you to span an NSS volume across multiple hard disk partitions. This is not desirable in a DRM configuration. Instead, maintain a one-to-one relationship among LUNs, remote copy sets, NSS partitions, NSS pools, and NSS logical volumes.

Red Hat Linux

Red Hat Linux operating systems EL Advanced Server V3.0 (32- and 64-bit) and SuSE SLES 8 (32- and 64-bit) support booting from the SAN.

Red Hat Linux supports bootless failover in standalone and in Lifekeeper cluster configurations that use Secure Path. Bootless failover using Qlogic Native Multipath is not supported because you cannot quiesce paths. With LifeKeeper clusters, generic and LVM resource kits for bootless DR group failovers are supported.

Sun

Sun Solaris does not support booting from the SAN or bootless failover. Sun uses UFS (UNIX file system). Sun cluster 3.1 is supported on Solaris V9 only.

SuSE Linux

SuSE Linux supports booting from the SAN on stand-alone systems only (not with clusters) for both 32- and 64-bit. SuSE SLES 8 and United Linux 1 support bootless failover in standalone and Lifekeeper cluster configurations that use Secure Path. Qlogic Native Multipath is not supported because you cannot quiesce paths. With LifeKeeper clusters, generic and LVM resource kits for bootless DR group failovers are supported. SuSE Linux uses file systems EXT2, EXT3, Reiser, and LVM.

Application design considerations

Some applications do not work well in a replication environment. Following are application design considerations:

- With a maximum of eight copy sets per application, some applications, such as Oracle or SAP, must be reconfigured to reduce the number of virtual disks to eight.
- Some applications, such as Microsoft Exchange, do not tolerate high average I/O latency and therefore may not be suitable for replication beyond a metropolitan area.

Supportable Solution Checklists

A

This appendix provides checklists to help ensure that you possess the required software and hardware for your Continuous Access EVA solution, excluding the storage area network (SAN). The checklist for the SAN is included in Appendix B.

Note: The following sections pertain to using a pair of arrays. When you use multiple replication relationships, adjust the rules to account for the odd array (for example, a third array at a third site). Each site should contain a copy of all pertinent documentation, installation kits, and copies of all necessary licenses.

Basic Continuous Access EVA configurations

Tables 20 through 28 contain a partial parts list for basic Continuous Access EVA configurations. Each table includes an "Order" column that you can use to track the amounts ordered. If you are upgrading a non-replicating EVA to a Continuous Access EVA, then reduce quantities by the amounts being reused.

It is helpful to use a drawing, like any of those in Figures 17 (on page 71), 21 (on page 76), 22 (on page 79), 24 (on page 84) and 25 (on page 85) to help you identify the parts and quantities you need for a complete solution.

EVA3000

For current information about the EVA3000, refer to the product website at http://h18006.www1.hp.com/products/storageworks/eva3000/. From this website, select the Specifications and warranty topic.

Table 20: EVA3000 virtual array storage system

Description	Part Number	Order
EVA3000 2C1D-C with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, one M5314 3U Dual-redundant FC loop 14-bay disk enclosure & appropriate mounting rails.	AD543A[1]	
EVA3000 2C2D-C/16x36 GB/15k with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 16x 36 GB 15K rpm HDDs.	344820-B21[1]	
EVA3000 2C2D-C 8x 72 GB/10k with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 8x 72 GB 10K rpm HDDs.	344821-B21[1]	
EVA3000 2C2D-C 8x 72 GB/15k with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 8x 72 GB 15K rpm HDDs.	344822-B21[1]	
EVA3000 2C2D-C/8x146 GB/10k with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 8x 146 GB 10K rpm HDDs.	344823-B21[1]	
EVA3000 2C2D-C/8x300 GB/10k with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 8x 300 GB 10K rpm HDDs.	AD535A[1]	
EVA3000 2C2D-C/8x250 GB with Foundation Service Solution One 3U Controller assembly with two HSV100 controllers w/redundant power supplies, two M5314 3U Dual-redundant FC loop 14-bay disk enclosures, appropriate mounting rails & 8x 250 GB HDDs.	AD517A[1]	
[1]0D1 displaying after the part number on your sales order indicates factory integration.		

Table 21: EVA3000 array options

Description	Part Number	Order
EVA Fibre Channel Cable Routing Spools (set of 12)	293357-B21	
EVA Rail-to-Rail Cable Kit for 1 G switches (set of 2)	293358-B21	
EVA Side-Rail Cable Kit for 2 G switches (set of 2)	293359-B21	
EVA3000 controller pair	344818-B21	
EVA3000 controller accessory kit	345438-B21	
EVA3000 M5314 drive enclosures	344819-B21[1]	
[1]0D1 displaying after the part number on your sales order indicates factory integration.		

Table 22: EVA3000 licenses

Order in pairs; licensing is based on the Continuous Access EVA usable replicated capacity, not on total array capacity, and is not dependent on Vraid type.

License Type	Part Number	Quantity	Order
base VCS license	T3589A	one per array	
1 TB replicated	331282-B21	one per array	
2 TB replicated	331284-B21	one per array	
4 TB replicated	331286-B21	one per array	
6 TB replicated	331287-B21	one per array	
8 TB replicated	331288-B21	one per array	
1 TB replicated—upgrade	344532-B21	one per array	
2 TB replicated—upgrade	331283-B21	one per array	
4 TB replicated—upgrade	331285-B21	one per array	

EVA5000

For current information about the EVA5000, refer to the product website at http://h18006.www1.hp.com/products/storageworks/enterprise/index.html. From this website, select the Specifications and warranty topic.

Table 23: EVA5000 virtual array storage system

Order in pairs and deploy one per site. Quantities are based on the storage system capacity and the number of required drives.

Description	Max Number of Drives per Storage System	60-Hz Part Number	50-Hz Part Number	Order
EVA5000 2C2D-C—42U graphite cabinet, 2 controllers, 2 disk enclosures	28	309620-B23	309620-B24	
EVA5000 2C6D-C—42U graphite cabinet, 2 controllers, 6 disk enclosures, 4 FC loop switches	84	283198-B23	283198-B24	
EVA5000 2C12D-C—42U graphite cabinet, 2 controllers, 12 disk enclosures, 4 FC loop switches	168	283199-B23	283199-B24	
EVA5000 8C8D-C—42U graphite cabinet, 8 controllers, 8 disk enclosures	112	283263-B23	283263-B24	
EVA5000 2C6D—42U Opal cabinet, 2 controllers, 6 disk enclosures, 3.1/ 6.2 TB (36/72-GB drives)[1]	84	234345-B21	234345-B22	
EVA5000 2C12D—42U Opal cabinet, 2 controllers, 12 disk drive enclosures, 6.2/12.3TB (36/72-GB drives)[2]	168	234346-B21	234346-B22	
EVA5000 0C6D Expansion—42U Opal cabinet, 6 disk enclosures, 8 FC interconnect cables to 2C12D, 8.7/17.5TB (36/72-GB drives)[1,2]	240 with 2C6D, 2C6D-A	234347-B21	234347-B22	
EVA5000 0C12D Expansion—42U Opal cabinet, 12 disk enclosures, 16 FC interconnect cables to 2C12D, 17.5/35TB (36/72-GB drives)[1,2]	240 with 2C6D, 2C6D-A	234348-B21	234348-B22	
EVA5000 2C6D-A—42U Opal cabinet, 2 controllers, 6 disk enclosures, four 12-port FC switches, 3.1/ 6.2 TB (36/72-GB drives)[1]	84	283194-B21	283194-B22	
EVA5000 2C12D-A—42U Opal cabinet, 2 controllers, 12 disk enclosures, four 12-port FC switches, 6.2/12.3TB (36/72-GB drives)[1]	168	283195-B21	283195-B22	

Table 23: EVA5000 virtual array storage system (continued)

Order in pairs and deploy one per site. Quantities are based on the storage system capacity and the number of required drives.

Description	Max Number of Drives per Storage System	60-Hz Part Number	50-Hz Part Number	Order
EVA5000 0C6D-A Expansion—42U Opal cabinet, 6 disk enclosures, 8 FC interconnect cables to 2C12D-A, 8.7/17.5TB (36/72-GB drives)[1]	240 with 2C6D, 2C6D-A	283196-B21	283196-B22	
EVA5000 0C12D-A Expansion—42U Opal cabinet, 12 disk enclosures, 16 FC interconnect cables to 2C12D-A, 17.5/35TB (36/72-GB drives)[1]	240 with 2C12D, 2C12D-A	283197-B21	283197-B22	
EVA5000 2C2D-B—41U Graphite cabinet, 2 controllers, 2 disk enclosures, 1.5/ 3.6 TB (36/72-GB drives)[2]	28	309620-B21	309620-B22	
EVA5000 2C6D-B—41U Graphite cabinet, 2 controllers, 6 disk enclosures, four 12-port FC switches, 3.1/ 6.2 TB (36/72-GB drives)	84	283198-B21	283198-B22	
EVA5000 2C12D-B—41U Graphite cabinet, 2 controllers, 12 disk drive enclosures, four 12-port FC switches, 6.2/12.3TB (36/72-GB drives)	168	283199-B21	283199-B22	
EVA5000 0C6D-B Expansion—41U Graphite cabinet, 6 disk enclosures, 12 FC interconnect cables to 2C12D-B, 8.7/17.5TB (36/72-GB drives)	240 with 2C6D, 2C6D-B	283264-B21	283264-B22	
EVA5000 0C12D-B Expansion—41U Graphite cabinet, 12 disk enclosures, 24 FC interconnect cables to 2C12D-B, 17.5/35TB (36/72-GB drives)	240 with 2C12D, 2C12D-B	283265-B21	283265-B22	
EVA5000 8C8D—421U Graphite or cabinet, 8 controllers, 16 Fibre Channel Loops, 8 disk enclosures, 4.1/8.2 TB (36/72 GB drives)[1]	28 (4 systems in one rack)	239782-001	249977-B21	
EVA5000 8C8D-B—41U Graphite or cabinet, 8 controllers, 16 Fibre Channel Loops, 8 disk enclosures, 4.1/8.2 TB (36/72 GB drives)	28 (4 systems in one rack)	283263-B21	283263-B22	

[1]Supported, but may not be orderable.

[2]Servers are not supported in these racks.

Table 24: EVA5000 array options

Description	Part Number	Order
Enterprise Utility Cabinet—42U Graphite	289191-B23 (60 Hz)	
	289191-B24 (50 Hz)	
Enterprise Utility Cabinet—41U Graphite	289191-B21	
EVA Disk Enclosure M5214-A for 41U	290475-B21	
EVA Disk Enclosure M5214 for 42U	232113-B21	
EVA Disk Enclosure M5314 for 42U	344819-B21	
EVA Fibre Channel Cable Routing Spools (set of 12)	293357-B21	
EVA Rail-to-Rail Cable Kit for 1 G switches (set of 2)	293358-B21	
EVA Side-Rail Cable Kit for 2 G switches (set of 2)	293359-B21	
EVA Backend Loop Switch Upgrade Kit*	309620-B21	
* HP strongly recommends the EVA backend loop switch upgrade when upgrading an existing 2C2D array to six or more drive enclosures.		

Table 25: EVA5000 licenses

Order in pairs; licensing is based on the Continuous Access EVA usable replicated capacity, not on total array capacity, and is not dependent on Vraid type.

License Type	Part Number	Quantity	Order
base VCS license	250203-B25	one per array	
VCS media kit V3.0c (required)	T3588A	one per array	
base Continuous Access user interface	331268-B21	one per appliance	
1 TB replicated	344530-B21	one per array	
2 TB replicated	331270-B21	one per array	
4 TB replicated	331272-B21	one per array	
6 TB replicated	331273-B21	one per array	
8 TB replicated	331274-B21	one per array	
10 TB replicated	331275-B21	one per array	
12 TB replicated	331276-B21	one per array	
14 TB replicated	331277-B21	one per array	
16 TB replicated	331278-B21	one per array	

Table 25: EVA5000 licenses (continued)

Order in pairs; licensing is based on the Continuous Access EVA usable replicated capacity, not on total array capacity, and is not dependent on Vraid type.

License Type	Part Number	Quantity	Order
18 TB replicated	331279-B21	one per array	
24 TB replicated	331280-B21	one per array	
36 TB replicated	331281-B21	one per array	
1 TB replication increment	344531-B21	one per array	
2 TB replication increment	331269-B21	one per array	
4 TB replication increment	331271-B21	one per array	

Table 26: Disk drives

Order in pairs and ship half to each site so that the total number of drives does not exceed 84, 112, or 168, according to the choice in Table 23 on page 100. Drives can be mixed and matched based on capacity needs. Multiple drive types may be installed within a single array. Disk groups, however, may contain only one drive type and a minimum of eight drives.

	Factory Installed—Part Number		Field Installed—	
Description	Non-EMEA	EMEA Only	Part Number	Order
300-GB 10,000-rpm, 2-Gbps, 1-inch drive	364622-B23	364622-B22	364622-B22	
250-GB FATA disk, 2-Gbps hybrid drive	364437-B23	364437-B22	364437-B22	
146-GB 10,000-rpm, 2-Gbps, 1-inch drive	293556-B23	293556-B22	293556-B22	
72-GB 15,000-rpm, 2-Gbps, 1-inch drive	293568-B23	293568-B22	293568-B22	
72-GB 10,000-rpm, 2-Gbps, 1-inch drive	238921-B21	238921-B22	238921-B22	
36-GB 15,000-rpm, 2-Gbps, 1-inch drive	236205-B23	236205-B22	236205-B22	
36-GB 10,000-rpm, 2-Gbps, 1-inch drive	238590-B23	238590-B22	238590-B22	

Required Software

Operating system solution platform kits are required. You can download kits at no cost and without part numbers at http://h18006.www1.hp.com/products/sanworks/softwaredrivers/securepath/index.html.

The following operating systems are supported:

- HP-UX
- Windows NT, Windows 2000, and Windows Server 2003 Enterprise Edition
- HP OpenVMS
- HP Tru64
- SUN Solaris
- IBM AIX
- NetWare
- Linux

Additional Required Software

For current information about Secure Path, refer to the website at http://h18006.www1.hp.com/products/sanworks/softwaredrivers/securepath/index.html.

Table 27: Additional required software for HP, Sun, Windows, IBM, Novel, or Linux

Order one per server, except HA/F500, which supports one pair of Microsoft Windows-based servers. Although this table does not list them, packages containing multiple Secure Path licenses are available.

Description	Part Number	Order
HP-UX, Secure Path Host, V3.0D	T3549A	
Sun Solaris, Secure Path Host, V3.0D	T3623A	
Windows Secure Path Host, V4.0C	165989-B25	
HA/F500 Enhanced DT (Disaster Tolerant) Kit for EVA	306791-B21	
Secure Path for IBM AIX, V2.0D	231495-B24	
Secure Path for Novell NetWare, V3.0C	165993-B22	
Secure Path for Red Hat Linux, V3.0C	T3575A	
Secure Path for SuSE Linux, V3.0C	T3575A	

Table 28: Fibre channel adapters

Order one pair per server for each supported single-channel adapter or one per server for each supported dual-channel adapter.

Operating System	Description	Part Number	Order
HP OpenVMS	1 Gbps PCI - SC	168794-B21*	
	FCA2354, 64-bit/66 MHz PCI, 2 Gbps LC	261329-B21	
	FCA2384, 64-bit/133 MHz PCI-X, 2 Gbps LC	302784-B21	
HP Tru64 UNIX	1 Gbps PCI - SC	168794-B21*	
	FCA2354, 64-bit/66 MHz PCI, 2 Gbps LC	261329-B21	
	FCA2384, 64-bit/133 MHz PCI-X, 2 Gbps LC	302784-B21	
	FCA2684, 64-bit/133 MHz PCI-X, 2 Gbps LC	DS-A5132-AA	
	FCA2684DC, 64-bit/133 MHz PCI-X, 2 Gbps Dual Channel	DS-A5134-AA	
HP Tru64 & OVM	FCA2384, PCI-X, 2 Gbps LC	302784-B21	
HP-UX	Native PCI, 1 Gbps SC	A5158A	
	Native PCI, 2 Gbps LC	A6795A	
	Native HP-HSC, 1 Gbps SC	A6685A	
	Native 64-bit PCI Dual Channel, 2 Gbps LC (HP-UX 11.11only)	A6826A	
	Native 64-bit PCI Dual Channel combo, 2 Gbps LC (HP-UX 11.11only)	A9782A	
	Native 64-bit PCI Dual Channel combo, 2 Gbps LC (HP-UX 11.11only)	A9784A	
IBM AIX	64-bit PCI, 1 Gbps SC	197819-B21	
	32-/64-bit, 66 MHz PCI, 2 Gbps	PC2000LC-HPSP	
MS Windows	PCI, 1 Gbps SC	176479-B21*	
	FCA2101, 64-bit/66 MHz PCI, 2 Gbps LC	245299-B21	
	FCA2355, PCI, dual port, 2 Gbps LC	308540-B21	
	FCA2214, PCI, 2 Gbps LC	281541-B21	
	FCA2214DC, PCI, dual port 2 Gbps LC	321835-B21	
	FCA2404, PCI-X, 2 GBPS LC	305573-B21	
	FCA2404DC, PCI-X, dual port, 2 Gbps LC	323264-B21	
	64-bit, 133 MHz, PCI-X, 2 Gbps	A7298A	

Table 28: Fibre channel adapters (continued)

Order one pair per server for each supported single-channel adapter or one per server for each supported dual-channel adapter.

Operating System	Description	Part Number	Order
Novell NetWare	FCA2210, PCI-X, 2 Gbps LC	281540-B21	
Red Hat Linux	FCA2214, PCI-X, 2 Gbps LC	281541-B21	
	FCA2214DC, dual port, PCI-X, 2 Gbps LC	321835-B21	
Sun Solaris	32-bit PCI, 1 Gbps SC	380576-001*	
	64-bit sBus, 1 Gbps SC, V2.6, 7, 8 only	123503-001*	
	FCA2257P, PCI, 2 gbps LC	254456-B21	
	FCA2257C, cPCI, dual port, 1 Gbps SC	254457-B21	
	FCA2257S, sBus, dual port, 1 Gbps SC	254458-B21	
SuSE Linux	FC2214, PCI-X, 2 Gbps LC	281541-B21	
	FCA2214DC, dual port, PCI-X, 2 Gbps LC	321835-B21	
* Supported. May not be orderable.			

SAN Solution Checklists

B

This appendix provides checklists to help ensure that you possess the required SAN components.

It is helpful to use a drawing, like any of those in Figures 17 (on page 71), 21 (on page 76), 22 (on page 79), 24 (on page 84) and 25 (on page 85) as an aid to identifying parts and quantities needed in a complete solution.

Table 29 lists B-series switches. Order in pairs, one per fabric at each site to have symmetric fabrics at each site, not necessarily between sites. For example, both fabrics at site 1 may contain only 1-Gbps switches, while both fabrics at site 2 may contain only 2-Gbps switches. Another option is five switches per fabric at one site and only three per fabric at the other. A maximum of 28 switches per fabric is supported. If mixing vendors, follow the limitations outlined in the *HP StorageWorks SAN Design Reference Guide*.

Table 29: B-series switches

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
Compaq StorageWorks SAN Switch 8[1]	158222-B21	8-port, (1-Gbps SC)	8	
HP Brocade 2400 (HP reseller)[2]				
Compaq StorageWorks SAN Switch 16[1]	158223-B21	16-port, (1-Gbps SC)	16	
HP Brocade 2800 (HP Reseller)[2]				
Compaq StorageWorks SAN Switch 8-EL[1]	176219-B21	8-port, (1-Gbps SC)	1	
Compaq StorageWorks SAN Switch 16-EL[1]	212776-B21	16-port, (1-Gbps SC)	16	
Compaq StorageWorks SAN Switch 2/16	287055-B21	16-port (2-Gbps, dual power LC)	16	
HP StorageWorks SAN Switch 2/16[2]				
Compaq StorageWorks SAN Switch 2/8-EL	258707-B21	8-port (2-Gbps, LC)	8	
HP Surestore FC 1Gb/2Gb Entry Switch 8B[2]				
HP Surestore FC 1Gb/2 Gb Switch 8B[2]				
HP StorageWorks SAN Switch 2/8-EL[2]				

Table 29: B-series switches (continued)

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
SAN Switch 2/8-EL, upgrade	288162-B21	2-Gbps, LC	8	
Compaq StorageWorks SAN Switch 2/16-EL	283056-B21	16-port (2-Gbps, LC)	16	
HP Surestore FC 1 Gb/2 Gb Switch 16B[2]				
SAN Switch 2/16-EL, upgrade	288250-B21	2-Gbps, LC	16	
HP StorageWorks Core Switch 2/64	254508-B21	32-port (2-Gbps, LC). Includes 32-port switch base unit, rack-mount kit, power cords, serial cable, and documentation	32	
StorageWorks Core Switch Blade 2/64	286352-B21	16-port (2-Gbps, LC)	16	
HP StorageWorks SAN Switch 2/32 Power Pak	311026-B21	2-Gbps, dual power LC	32	
Compaq StorageWorks SAN Switch Integrated/32		Contains six fully interconnected SAN switch 16, 32-port SC	0	
HP Surestore FC Switch 6164 (64 ISL Ports)[2]				
Compaq StorageWorks SAN Switch Integrated/64		Contains six fully interconnected SAN switch 16, 64-port SC	0	
HP Surestore FC Switch 6164 (32 ISL Ports)[2]				
[1]Supported; may not be orderable. [2]Identical device and may be substituted.				

Table 30: B-series switch options

Option	Applicable To	Option Part Number	Order
Remote switch	SAN Switch 2/8 and 2/16	325891-B21	
Remote switch	SAN Switch 2/31	313460-B21	
Remote switch	SAN Switch 2/64	330883-B21	
HP StorageWorks Redundant Power Supply	SAN Switch 2/16	260255-B21	
HP StorageWorks Fabric Manager software	SAN Switch 2/32	313455-B21	
HP StorageWorks ISL Trunking	SAN Switch 2/64	288144-B21	
HP StorageWorks Extended Fabric	SAN Switch 2/64	288141-B21	

Table 30: B-series switch options (continued)

Option	Applicable To	Option Part Number	Order
FC optical shortwave GBIC (1-Gb SC)	SAN Switch 8, 16	380561-B21	
FC connect kit, 3-SW GBICs, 2-2m cables (SC to SC)	SAN Switch 8, 16	380579-B21	
FC connect kit, 2-SW GBICs, 2-2m cables (SC to SC)	SAN Switch 8, 16	380596-B21	
HP StorageWorks Fabric Watch	SAN Switch 2/16	262864-B21	
HP StorageWorks ISL Trunking	SAN Switch 2/16	262872-B21	
HP StorageWorks Advanced Perf Monitor	SAN Switch 2/16	262870-B21	
HP StorageWorks Performance Bundle*	SAN Switch 2/16	262874-B21	
HP StorageWorks QuickLoop	SAN Switch 2/16	262866-B21	
HP StorageWorks Extended Fabric	SAN Switch 2/16	262868-B21	
FC SFP Shortwave Transceiver Kit (2-Gbps, LC)	All SAN Switch 2 models	221470-B21	
10 km Longwave Transceiver (2-Gbps, LC)	All SAN Switch 2 models	300835-B21	
35 km Longwave Transceiver (2-Gbps, LC)	All SAN Switch 2 models	300836-B21	
* Includes Fabric Watch, ISL Trunking, and Performance Monitor software.			

Table 31 lists C-series switches. HP resells two classes of Cisco switches: the MDS 9509/9506 Multilayer Director Switch and the MDS 9216 Multilayer Fabric Switch. Both switches allow for the addition of multiple module options, including 16- and 32-port 1- and 2-Gb Fibre Channel switching modules.

A total of 7 of the MDS 9000 16- or 32-port FC modules is supported in one base unit.

Table 31: C-series switches

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
MDS 9120	346700-B21	20-port fabric switch	20	
MDS 9140	346701-B21	40-port fabric switch	40	
MDS 9509 Multilayer Director	332306-B21	Base unit with 0 ports	N/A	
MDS 9506 Multilayer Director	346702-B21	Base unit with 0 ports	N/A	

Table 31: C-series switches (continued)

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
MDS 9000	332307-B21	16-port 1 / 2-Gpbs FC module	16	
MDS 9000	332308-B21	32-port 1 / 2-Gpbs FC module	32	
MDS 9216	332315-B21	16-port 2-Gpbs FC + 1-slot	16	

Table 32: C-series switch options

Option	Applicable to	Option Part Number	Order
2500W DC power supply	MDS 9509	332309-B21	
1900W AC power supply	MDS 9506	346704-B21	
1900W DC power supply	MDS 9506	346705-B21	
2500W AC power supply	MDS 9509	333739-B21	
4000W AC power supply	MDS 9509, US	332310-B21	
4000W AC power supply	MDS 9509, Int'l	332311-B21	
MDS 9500 8-port 1GbE IP storage module		332907-B21	
MDS 9500 sup compact flash disk, 512 MB		332314-B21	
MDS 9000 port analyzer adapter		332317-B21	
MDS 9000 IP services SW license		346703-001	
MDS 9200 IP services SW license		348232-001	
1470 NM CWDM FC SFP		347419-B21	
1490 NM CWDM FC SFP		347420-B21	
1510 NM CWDM FC SFP		347421-B21	
1530 NM CWDM FC SFP		347422-B21	
1550 NM CWDM FC SFP		347423-B21	
1570 NM CWDM FC SFP		347424-B21	
1590 NM CWDM FC SFP		347425-B21	
1610 NM CWDM FC SFP		347426-B21	
4 wavelength add/drop mux		347427-B21	

Table 32: C-series switch options

Option	Applicable to	Option Part Number	Order
8 wavelength mux/demux		347428-B21	
Fibre Channel LC/SC 1 meter		347429-B21	
Fibre Channel LC/SC 5 meter		347430-B21	
1 Gbps Ethernet, 2 Gbps Fibre Channel-SW SFP, LC		336223-B21	
1 Gbps Ethernet, 2 Gbps Fibre Channel-LW SFP, LC		336224-B21	
2 Gbps Fibre Channel-SW SFP, LC		332312-B21	
2 Gbps Fibre Channel-LW SFP, LC		332313-B21	

Table 33 lists M-series switches. Order in pairs, one per fabric at each site to have symmetric fabrics at each site, not necessarily between sites. For example, both fabrics at site 1 may contain only 1-Gbps switches, while both fabrics at site 2 may contain only 2-Gbps switches. Another option is five switches per fabric at one site and only three per fabric at the other. A maximum of 24 switches per fabric is supported. If mixing vendors, follow the limitations outlined in the *HP StorageWorks SAN Design Reference Guide*.

Table 33: M-series switches

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
HP StorageWorks Edge Switch 2/16[1]	286811-B21	16-port (2-Gbps, dual power, LC). Includes 16 ports and 16 shortwave transceivers.	0	
HP StorageWorks Edge Switch 2/24	316095-B21	16-port (2-Gbps, dual power, LC)	0	
HP StorageWorks Edge Switch 2/24	316096-B21	8-flexport upgrade. Includes 8 ports and 8 shortwave transceivers.	0	
HP StorageWorks Edge Switch 2/32	286810-B21	8-port (2-Gbps, dual power, LC). Includes 8 ports and 8 shortwave transceivers; additional ports may be added in groups of 8 using 8-flexport upgrade.	0	

Table 33: M-series switches (continued)

Name	Part Number	Description	Number of GBIC/SFPs Needed	Order
HP StorageWorks Edge 2/32	302660-B21	8-flexport upgrade. Includes 8 ports and 8 shortwave transceivers.	0	
HP StorageWorks Director 2/64[1]	286809-B21	32-port (2-Gbps, LC). Comes standard with 32 shortwave ports and can support up to 64 ports; upgradable in 8-port increments.	0	
HP StorageWorks SAN Director 64	254512-B21			
HP Surestore Director FC-64[1]	A6534A			
SAN Director 2/64 2/140, 8-port module kit	300833-B21	8 ports, 8SFPs, 2-Gbps, LC	0	
HP StorageWorks Director 2/140	316093-B21	64-port (2-Gbps, LC). Comes standard with 64 shortwave ports and can support up to 140 ports; upgradable in 4- and 8-port increments.	0	
SAN Director 2/64 and 2/140 Shortwave 4-port Module Kit	316094-B21	4 ports, 4 SFPs, 2-Gbps, LC	0	
SAN Director 2/140 Shortwave 8-port Module Kit	300833-B21	8-ports, 8 SFPs, 2-Gbps, LC	0	

Table 34: M-series switch options

Option	Applicable to	Option Part Number	Order
HP StorageWorks Edge Switch Product Manager License[1]	SAN Switch 2/16	300659-B21	
HP StorageWorks Edge Switch SANtegrity Binding OS*	SAN Switch 2/16	317069-B21	
HP StorageWorks Edge Switch Product Manager License	SAN Switch 2/24	317067-B21	
HP StorageWorks M-series Fabric Manager Software	Edge Switch	287406-B22	
HP StorageWorks Edge Switch SANtegrity Binding OS*	SAN Switch 2/24	317070-B21	
HP StorageWorks Edge Switch Product Manager License	SAN Switch 2/32	300658-B21	

Table 34: M-series switch options (continued)

Option	Applicable to	Option Part Number	Order
HP StorageWorks Edge Switch SANtegrity Binding OS[1]	SAN Switch 2/32	317071-B21	
HP StorageWorks SAN Director Omnibook add-on kit[2]	Fabric	303003-B21	
HP StorageWorks SAN Director M-series Fabric Manager[3]	Fabric	287406-B21	
SAN Director 2/64, 2/140 Rack Mount Kit	M-series racks; Compaq 9000 and 10000 racks	254520-B21	
SAN Director 2/64 Rack Mount Kit	M-series racks; Compaq 9000 and 10000 racks	302659-B21	
HP StorageWorks Director Switch SANtegrity Binding OS[1]	SAN Switch 2/64	317072-B21	
SAN Director Shortwave Optical Transceiver Kit, 2-Gbps, LC	SAN Director 2/64 and 2/140	300834-B21	
10 km Long-Distance Transceiver Kit, 2-Gbps, LC	SAN Director 2/64 and 2/140	300835-B21	
35 km Long-Distance Transceiver Kit, 2-Gbps, LC	SAN Director 2/64 and 2/140	300836-B21	
SAN Director Shortwave Port Module Kit, 1-Gbps, LC	SAN Director 2/64 and 2/140	254516-B21	
SAN Director Longwave Port Module Kit, 1-Gbps, LC	SAN Director 2/64 and 2/140	254515-B21	
SAN Director Combination Shortwave and Longwave Port Module Kit, 1-Gbps, LC	SAN Director 2/64 and 2/140	254517-B21	

[1] SANtegrity Binding Security software locks down the fabric, preventing intentional security threats and limiting exposure to accidental changes to the storage network.
[2] Select one license per fabric.
[3] Select one license per fabric and select the appropriate Product Manager licenses for each additional switch in the SAN.

Table 35: Fibre Channel cables GBIC/SFP MMF cables

Order one per FCA port, four per EVA, one per ISL, and one per gateway.

Description	Part Number	Order
SC-SC coupler, used to join two SC-SC or SC-LC cables together	C7534A	
SC-SC Fibre Channel Cable Kit (2 GBICs, two 2-m cables)	380596-B21	
SC-SC Fibre Channel cable, multimode (2-m)	234457-B21	
SC-SC Fibre Channel cable, multimode (5-m)	234457-B22	
SC-SC Fibre Channel cable, multimode (15-m)	234457-B23	
SC-SC Fibre Channel cable, multimode (30-m)	234457-B24	
SC-SC Fibre Channel cable, multimode (50-m)	234457-B25	
LC-SC Fibre Channel cable, optical (0.6-m)	221691-B25	
LC-SC Fibre Channel cable, optical (1-m)	221691-B24	
LC-SC Fibre Channel cable, optical (2-m)	221691-B21	
LC-SC Fibre Channel cable, optical (5-m)	221691-B22	
LC-SC Fibre Channel cable, optical (15-m)	221691-B23	
LC-SC Fibre Channel cable, optical (30-m)	221691-B26	
LC-SC Fibre Channel cable, optical (50-m)	221691-B27	
LC-LC Fibre Channel cable, multimode (0.6-m)	221692-B25	
LC-LC Fibre Channel cable, multimode (1-m)	221692-B24	
LC-LC Fibre Channel cable, multimode (2-m)	221692-B21	
LC-LC Fibre Channel cable, multimode (5-m)	221692-B22	
LC-LC Fibre Channel cable, multimode (15-m)	221692-B23	
LC-LC Fibre Channel cable, multimode (30-m)	221692-B26	
LC-LC Fibre Channel cable, multimode (50-m)	221692-B27	

Long-wave SMF cable, ordered from a third-party vendor, can be used for direct switch-to-switch connections up to 100 km at 1 Gbps, or 35 km at 2 Gbps. The cable must be duplex tight-buffered SMF 9/125 µm (Belcore GR-409 or ITU G.652) compliant. The connectors must be SC duplex low-metal (NTT-SC Belcore 326 or IEC-874-19 SC) compliant.

The same long-wave SMF cable is also used for most WDM applications. Contact your WDM vendor for exact requirements.

Storage Management Appliances

You must order at least two SMAs, one per site. HP recommends four SMAs, two per site, so that in the event of a disaster, there is a standby appliance available at each site.

Table 36: Storage Management Appliances

Description	Part Number	Order
HP StorageWorks Management Appliance II	189715-002	
HP StorageWorks Management Appliance I (supported but not orderable)	N/A	

Continuous Access EVA-over-WDM or EVA-over-very-long-distance direct-fiber configurations

If you are using a Continuous Access EVA-over-WDM or an EVA-over-a-very-long-distance direct-fiber configuration, and are using B-series switches, you must order the item in Table 37 in addition to the items in Table 17 through Table 28.

Table 37: Other required items for long-distance direct fiber connections

Order MMF cables and GBICs to connect a WDM multiplexer to a switch.

Description	Quantity Required	Part Number	Order
Extended Fabric License for B-series Switches	4 (two per site)	262868-B21	

Additional considerations (order items from third-party vendors):

- Each WDM link supports full bandwidth Fibre Channel connections up to 1 Gbps when using WDM.
- WDM uses the same 9-µm cable for the intersite link. WDM may be used for multiple site-to-site links, such as the network, in addition to Fibre Channel connections.
- Only one switch may be connected directly to the WDM interface.

Continuous Access EVA Fibre Channel-over-IP configurations

If you are using a Continuous Access EVA-over-IP configuration, you will require the items in Table 38 in addition to the items in Table 17 through Table 28.

Table 38: Additional items for Fibre Channel-over-IP

Description	Quantity Required	Part Number	Order
Remote switch for 2/8 and 2/16	4 (two per site)	325891-B21	
Remote switch for 2/32	4 (two per site)	313460-B21	
Remote switch for core 2/64	4 (two per site)	330883-B21	
Fibre Channel-to-IP Gateway	4 (two per site)	order from vendor	

Reference

C

This appendix provides details about switches, FCAs, and maximum fiber cable distances.

Table 39: Maximum distances over fiber

Type of Fiber Cable	Maximum Distance at 1 Gbps	Maximum Distance at 2 Gbps
50-mm multimode	500 m	300 m
62.5-mm multimode fiber	200 m	150 m
9-mm (VLD GBICs)	100 km	35 km
9-mm (LD GBICs)	10 km	10 km

Use Table 40 to cross reference capabilities such as interface and performance. Only the HP version is supported due to uniqueness in the default settings, firmware, and drivers.

Table 40: FCA model number cross-reference to vendor model number

HP Model Number	Vendor Model Number[1]	Description	Connector Type	Operating System
FCA2101	Emulex LP952L	PCI, 64-bit, 66-MHz, 2-Gbps	LC	Windows
KGPSA-CB	Emulex LP8000	PCI, 64-bit, 33-MHz, 1-Gbps	SC	Windows
FCA2257C	QLogic QCP2202F/33	cPCI, DC, 64-bit, 33-MHz, 1-Gbps	SC	Solaris
FCA2257P	QLogic QLA2310F	PCI-X, 64-bit, 66-MHz, 2-Gbps	LC	Solaris
FCA2257S	QLogic QLA2202FS	S-bus, DC, 64-bit, 25-MHz, 1-Gbps	SC	Solaris

Table 40: FCA model number cross-reference to vendor model number (continued)

HP Model Number	Vendor Model Number[1]	Description	Connector Type	Operating System
FCA2355	Emulex LP9002DC	PCI, 64-bit, 66-MHz, 2-Gbps, dual port	LC	OpenVMS, Tru64
FCA2354	Emulex LP9002	PCI, 64-bit, 66-MHz, 2-Gbps	LC	OpenVMS, Tru64
FCA2384	Emulex LP9802	PCI-X, 64-bit, 133-MHz, 2-Gbps	LC	OpenVMS, Tru64
FCA2404	Emulex LP9802	PCI-X, 64-bit, 133-MHz, 2-Gbps	LC	MS Windows
FCA2404DC	Emulex LP9802DC	PCI-X, 64-bit, 133-MHz, 2-Gbps, dual port	LC	MS Windows

[1] The vendor model number is provided for reference to capabilities common to all versions of that adapter.

Glossary

This glossary defines terms used in this guide or related to Continuous Access EVA and is not a comprehensive glossary of computer terms.

actual disk failure protection level
The actual level of protection available to the disk group as specified in its current configuration.

allocation policy
Storage system rules that govern how virtual disks are created. There are two rules:

- Allocate Completely—The space a virtual disk requires on the physical disks is reserved, even if the virtual disk is not currently using the space.
- Allocate on Demand—The space a virtual disk requires on the physical disks is not reserved until needed.

alternate site
See remote site.

asynchronous replication mode
The mode of operation of a data replication (DR) group whereby the EVA controller firmware provides I/O completion to the server after data is delivered to cache at the source, and before the data delivery to cache at the destination. *See also* DR group write mode.

asynchronous transfer mode (ATM)
Communications networking technology for LANs and WANs that carries information in fixed-size cells of 53 bytes (5 protocol and 48 data).

bandwidth
The transmission capacity of a link or system, usually measured in bits per second.

bandwidth latency product
The measurement of the ability to buffer data and is the raw transfer speed in bytes/sec x the round-trip latency in seconds.

bidirectional
The circumstance when a storage system is configured so that it contains both source and destination virtual disks.

bit error rate (BER)
The rate at which a single bit error is detected in communications networks.

B-series switches
Fibre Channel core and SAN switches made by Brocade and sold by HP. See Table 28 on page 105 for a list of Continuous Access EVA-supported devices of this type.

CAC
Corrective action code. A display component that defines the action required to correct a problem. This component is displayed on the HP StorageWorks Command View EVA graphical user interface (GUI).

Continuous Access
A storage–based HP StorageWorks solution consisting of two storage systems performing storage array–to–storage array replication, along with the management user interface (Continuous Access UI) that facilitates configuring, monitoring, and maintaining the replicating capabilities of the storage systems.

CGR
See constant bit rate.

clone
A physical, block-for-block copy of an entire volume. The cloning capabilities reside in the HSV controller. The cloning operation creates a third copy of the parent volume. After the clone is created, it acts as a local mirror until data on the clone has been normalized and the clone broken off. After the clone is broken off, it can be mounted on another server that is connected to the controller.

Command View EVA
The Command View EVA consists of:

- The graphical user interface (GUI) that displays the usage of the storage system.
- The software behind the GUI, which controls and monitors the functions.

The Command View software can be installed on more than one HP OpenView Storage Management Appliance in a fabric. Each installation of the Command View software is a management agent. The client for the agent is a standard browser.

console LUN
A SCSI-3 virtual object that makes a storage array accessible by the server before any virtual disks are created. Also called a *communication LUN*.

console LUN ID
The ID that can be assigned when a server operating system requires a unique ID. The console LUN ID is assigned by the user, usually when the storage system is initialized. *See also* console LUN.

constant bit rate (CBR)
Category of ATM Quality of Service (QoS) that supports a constant or guaranteed data rate. CBR supports applications that need a highly predictable transmission rate.

Continuous Access UI
A tool for managing the replication of storage objects in a SAN. It provides a graphical user interface for the management of disk I/O, failover, and maintenance operations.

controller heartbeat LED
A green LED that flashes on the controller OCP to indicate that the HSV controller is operational.

copy set
Generic term for a logical disk in one storage array that is replicated to another logical disk in another storage array. There are two states: *normal* and *copying*. The term is commonly used to represent the pair of virtual disks, one on the source array and one on the destination array.

corrective action code
See CAC.

C-series switches
Switches made by Cisco. See Table 31 on page 109 for a list of devices of this type.

data distribution
Pushing copies of data to geographic locations to make it more easily accessible to many customers.

data entry mode
The state in which controller information can be displayed or controller configuration data can be entered. On the EVA storage system, the controller data entry mode is active when the LCD on the HSV controller OCP is flashing.

data migration
Moving data to a new location or to a logical disk with a different capacity.

data mining
A process that makes data available so that undiscovered and useful information can be extracted, analyzed, or tested.

data movement
Activities such as data backup, data migration, and data distribution.

data replication mode

See DR (data replication) mode.

default disk group

The first disk group created when the storage system is initialized. The default disk group can contain up to the entire set of physical disks in the array. The minimum number of physical disks the default disk group can contain is eight. The maximum is the number of installed disks and is limited to 14 drives per attached drive enclosure.

destination

Used as a context-sensitive adjective, destination describes the targeted recipient (DR group, array/EVA, storage system, and so on) of an I/O after it is replicated.

destination Vdisk

A virtual disk that is the recipient of replicated data from a source virtual disk.

destination Vdisk access

A storage system's ability to allow server access to a destination virtual disk. There are two options:

- Disabled
- Read-Only (write-protected; for future use)

disaster recovery (DR)

The ability to respond to an interruption in services by implementing an in-place plan for restoring an organization's critical business functions.

disaster tolerance (DT)

The capability for rapid recovery of user data from a remote location when a significant event or disaster occurs at the local computing site. It is a special combination of high-availability technology and services that can continue the operation of critical applications in the event of a site disaster. DT systems are designed to allow applications to continue operating during the disaster recovery period.

disk failure protection

For each disk group, the controllers hold in reserve the space in the physical disk pool equivalent to a selected number of physical disk drives.

The three levels of disk failure protection are:

- None—No protection is present.
- Single— The capacity of one physical disk is reserved.
- Double—The capacity of two physical disks is reserved.

Disk failure protection occurs when the storage array sets aside reserved capacity to take over the functionality of a failed or failing physical disk drive. In groups with mixed capacity disk drives, the reserved capacity is based on the largest disk in the disk group. The system must cover a failure in any drive, so it reserves enough capacity to cover the largest failure that could happen.

disk group

A named group of physical disks selected from all the available physical disks in the storage system where one or more virtual disks can be created. A physical disk can belong to only one disk group.

disk group occupancy alarm level

An event code that is generated when the amount of data stored in the disk group reaches a peak level of the total disk group capacity. For example, if the disk group capacity is 288 GB, and the occupancy alarm level is 80 percent, the event code is generated when the amount of data in the disk group reaches 230.4 GB. The default occupancy alarm level is 95 percent (274 GB in this example) of the total disk group capacity.

disk migration state

The state of data on the physical disk drive. Two states are possible:

- Stable—The physical disk drive has not failed.
- Migrating—The disk drive is failing or failure is predicted. If this state occurs, data is moved by the controller from the failing disk to other disk drives in the same disk group.

disk replacement delay

The time that elapses between detection of a possible drive failure and when the controller starts searching for spare disk space. Drive replacement seldom starts immediately, in case the failure was a glitch or temporary condition.

DR (data replication) group
A VCS construct organizing one or more virtual disks in an HSV storage system so that they replicate to the same specified destination, failover together if a single virtual disk in the collection fails, and preserve write ordering within the collection. This is the HSV implementation of an association set, and its member virtual disks are the HSV implementation of copy sets. Note that DR is used in context to mean either disaster recovery or data replication.

DR (data replication) group direction
The replication direction of a DR group. There are two states:

- Original (from *Home*)
- Reversed (failed over, or toward *Home*)

DR (data replication) group log state
The current behavior of the log associated with a DR group. In the state options, references to multiple destinations are for future use. There are three possible states:

- Idle—No destination is logging or merging.
- Logging—At least one destination is logging; none are merging.
- Merging—At least one destination is merging.

DR (data replication) group write mode
Characterizes how a write from a server is replicated. A DR group has two modes of replicating writes from the source virtual disk to the destination virtual disk:

- Synchronous Replication—A server write is acknowledged as complete by the storage array after the write command and data are inserted into the source and destination storage array cache. If the destination array is not available or the DR group is merging, the write command and data are inserted into the tail of the write history log.
- Asynchronous Replication—A server write is acknowledged as complete by the storage array after the write command is inserted into the source array cache.

DR (data replication) mode
The operational mode of a DR group that indicates the capability of I/O to be written to its source or its destination, or to both. There are four options:

- Active/Passive (source)
- Active/Passive (destination)
- Active/Active (master)
- Active/Active (non-master)

Both Active/Active modes are for future use.

dual fabric
Two independent fabrics providing multipath connections between FC end devices.

E1
The standard European carrier for transmission at 2.048 Mbps.

E2
The standard European carrier for transmission at 8.192 Mbps.

E3
The standard European carrier for transmission at 34.368 Mbps.

E4
The standard European carrier for transmission at 139.264 Mbps.

E5
The standard European carrier for transmission at 565 Mbps.

event
Any change that is significant to the storage system. Events include:

- A state change in hardware
- A state change in a logical element, such as a virtual disk
- A completion of a procedure
- An environmental change
- An operational failure
- A change in configuration, such as a new virtual disk that has been created or a new physical disk has been inserted

FATA (Fibre Attached Technology Adapted)
High performance disk drives that are used for data reference, data archive, data replication, and applications that use files infrequently.

fabric
A network of at least one Fibre Channel switch and attached devices.

failover
This term is context specific.

- **DR (data replication) group failover**—An operation to reverse the direction of a DR group.
- **Managed set failover**—An operation to reverse the direction of all the DR groups in the managed set.
- **Fabric or path failover**—The act of transferring I/O operations from one fabric or path to another.
- **Controller failover**—When a controller assumes the workload of its partner.

failsafe mode
When enabled, this is a data replication group mode in which all source virtual disks in the group become both unreadable and unwritable if any of their corresponding member virtual disks is unreachable. No write history logging takes place for the virtual disks. There are two states: *enabled* and *disabled*.

failsafe-locked
A condition that prohibits any further server I/O to any of the virtual disks in the data replication (DR) group. A failsafe-enabled DR group becomes failsafe-locked whenever it is unable to complete a write to its destination, or when a member virtual disk fails. It requires immediate intervention.

FC connection
A Fibre-Channel path between two storage systems or a server and its storage. A connection is made up of multiple FC links.

FC link
Fibre-Channel link. A path between two adjacent FC ports.

FCIP (Fibre Channel over Internet Protocol)
A method of moving Fibre Channel data over a standard Internet Protocol network. The IP network may be Ethernet, SONET, or ATM based.

Fibre Channel (FC)
Technology for very high speed, switching-based serial transmissions.

Fibre Channel connection
See FC connection.

flush cache
The act of writing data from cache to storage media.

full copy

A copy operation for which all 1-MB blocks written on a source virtual disk since it was created are replicated to a destination virtual disk.

Gbps or Gb/sec

Gigabits per second. A measurement of the rate at which the transfer of bits of data occurs. Nominally, 1 Gbps is a transfer rate of 1,000,000,000 (10^9) bits per second.

For Fibre Channel transceivers or FC loops the Gb transfer rates are:

- 1 Gbps is a transmission rate of 1,062,500,000 bits per second.
- 2 Gbps is a transmission rate of 2,125,000,000 bits per second.

gigabit interface converter (GBIC)

The hardware devices inserted into the ports of the Fibre Channel switch that hold the fiber optic cables in 1-Gb fabrics. GBIC devices are available for short-range applications (0.5 to 500 m), long-range applications (up to 10 km), and very long distances (up to 100 km).

gigabit link module (GLM)

Permanently installed GLMs to provide fiber optic cable transmission at distances of 0.5 to 500 m, depending on the cable size and quality of its installation.

heterogeneous SAN support

The ability for the product to operate with different operating systems and storage systems in a SAN.

hierarchical storage virtualization (HSV)

A transparent abstraction of storage at the block level. Using block-level mapping techniques, storage virtualization presents servers with a logical view of storage in the form of virtual disks. All raw storage is pooled, and virtual disks (or virtual disks) are created that draw their capacity from the pool.

high availability (HA)

Redundant systems, software, and IT processes to reduce the risk of downtime. No single points of failure.

home

The preferred storage system for the source virtual disks of a data replication group. By default, home is the storage system on which a source is created, although this designation is user settable.

homogeneous SAN support

Implies the ability for the product to operate with homogeneous operating systems and homogeneous storage systems in a SAN.

hop
One interswitch link.

HSV
See hierarchical storage virtualization.

I/O module
Input/output module. The enclosure element that is the FC-AL interface to the server or controller. I/O modules are bus speed specific: either 1 Gbps or 2 Gbps.

IDX
A 2-digit decimal number portion of the HSV controller termination code display that defines one of 32 locations in the termination code array that contains information about a specific event.

in-band communication
Communication that uses the same communications link as the operational data. *See also* out-of-band communication.

initialization
A configuration step that binds the controllers together as a storage array and establishes preliminary data structures on the disk array. Initialization also sets up the first disk group, called the default disk group, and makes the storage system ready for use.

IOPS
I/O operations per second.

ISL
Intersite link or interswitch link. The abbreviation is context sensitive.

license key
A license key is required to operate the HSV controller software. The license key is a WWN-encoded sequence that is obtained from the HP license key fulfillment website. Two types of license keys exist:

- Business Copy EVA— This license key is needed to unlock the snapshot and Snapclone features. This license can be added any time *after* the system has been initialized.
- Continuous Access EVA—This license is necessary to activate the data replication features of the HSV controller software. It is also necessary for running the Continuous Access UI.

link
A connection between two adjacent Fibre Channel ports, consisting of a pair of simplex fibers—one for each direction of information flow. An example is the connection between the Fibre Channel switch port and the HSV controller.

local site
One or more storage arrays residing at the local data center, which contains the source data. A storage system may be considered a local site for some virtual disks and a remote site for other virtual disks.

log
Storage that is used for logging.

logging
Usually context sensitive. In Continuous Access EVA, *logging* refers to the history of server write commands (and data) when the destination array is not accessible. If failsafe mode is enabled, logging does not occur.

logical unit number
See LUN.

long-wave GBIC
A GBIC used with 9-µm single mode fiber and operating on a wavelength of 1350 nm. The 1-Gbps model has SC-style connectors and runs to a distance of 10 km. The 2-Gbps SFP has LC-style connectors and runs to a distance of 30 to 40 km. Also known as an *LD* (long-distance) *GBIC*.

LUN
Logical unit number. An identifier through which a virtual disk is presented to a server.

managed set
Any collection of data replication (DR) groups selected by the user for the purpose of managing them. For example, a managed set can be created to manage all DR groups whose sources reside in the same cabinet or all DR groups that deal with a particular set of applications.

master
The controller of the storage array that powers up first. Also called *storage system master*.

Mbps
Megabits per second. Typically, a unit of measurement for bandwidth of a serial link.

MBps
Megabytes per second. Typically, a unit of measurement for throughput for a serial link.

merge
See merging.

merging
Transferring the contents of the write history log to the destination virtual disk to synchronize the source and destination virtual disks.

MMF
Multimode fiber, typically 50-µm, although 62.5-µm is also supported at reduced distances.

M-series switches
Fibre Channel Director and Edge switches made by McDATA and sold by HP. See Table 33 on page 111 for a list of Continuous Access EVA-supported devices of this type.

µm
Micrometer (10^{-6} meter).

nm
Nanometer (10^{-9} meter).

near-online storage
On-site storage of data on media that takes only slightly longer to access than online storage kept on high-speed disk drives.

OC1
The optical carrier signaling rate synchronous data handling (SDH) (STM0)/SONET of 51.84 Mbps.

OC3
The optical carrier signaling rate synchronous data handling (SDH) (STM0)/SONET of 155.52 Mbps.

OC12
The optical carrier signaling rate synchronous data handling (SDH) (STM0)/SONET of 622.08 Mbps.

OC48
The optical carrier signaling rate synchronous data handling (SDH) (STM0)/SONET of 2488.32 Mbps.

OCP
Operator control panel. The element on the front of an HSV controller that displays the controller's status using LEDs and an LCD. Information selection and data entry are controlled by the OCP onscreen buttons.

online storage
An allotment of storage space that is available for immediate use, such as a peripheral device that is turned on and connected to a server.

operation state

Current operating condition of a system component. There are three states:

- Normal
- Failed
- Attention (indicates possible problem)

out-of-band communication

Communication that uses a different communications link than that used by operational data.

See also in-band communication.

peak cell rate (PCR)

Peak cell rate is the maximum transmission speed of a virtual connection and is a required parameter for the CBR Quality of Service (QoS) category.

permanent virtual circuit (PVC)

Logical connection between two points that are manually defined by the network administrator.

preferred path

A preference for which controller of the storage array manages the virtual disk. This preference is set by the user through Command View EVA when creating the virtual disk. The primary purpose of preferring a path is load balancing.

presentation

The process whereby a controller presents a virtual disk only to the server that has authorized access.

primary site

See local site.

quality of service (QoS)

Each virtual connection in a communications network has a service category. The performance of the connection is measured by the established QoS parameter.

reconstruction

The process of regenerating the contents of a failed member data. The reconstruction process writes the data to a spare set disk and incorporates the spare set disk into the mirrorset, striped mirrorset, or RAIDset from which the failed member came.

relationship

The arrangement created when two storage systems are partnered for the purpose of replicating data between them.

remote site

A site containing one or more storage arrays that contains copies of data stored at a local site. A remote site could be in the same room or building as the local site, but usually it is not.

reservation state

Three possible reservation states exist for a virtual disk:

- None—No server has a reservation on the virtual disk.
- Regular—One server has a reservation on the virtual disk. A regular reservation will not be preserved through failovers or power failures.
- Persistent—One or more servers have the virtual disk reserved. A persistent reservation is normally preserved through failovers and power failures. A persistent reservation can be released by other servers.

resume

Command issued to a data replication (DR) group or managed set that causes replication to resume after being suspended. This command may initiate a merging of the DR group log or a full copy. *See also* suspend.

SCSI (Small Computer System Interface)

An American National Standards Institute (ANSI) interface standard defining the physical and electrical parameters of a parallel I/O bus. A processor-independent standard protocol for system-level interfacing between a computer and intelligent devices, including hard drives, disks, CD-ROM drives, printers, scanners, and other devices.

SFP (small form factor pluggable GBIC)

A 2-Gbps GBIC. Typically uses an LC connection.

shortwave GBIC

A GBIC used with 50- or 62.5-µm multimode fiber and operating on a wavelength of 850 nm. The 1-Gbps model has SC-style connectors and runs to a maximum of 500 m. The 2-Gbps SFP has LC-style connectors and runs to a maximum of 300 m. Also known as an *SD* (short distance) *GBIC*.

single path

A single connection or path between storage systems containing source and replicated data, or between a server and the storage assigned to that server.

site failover

Command to change a destination role to a source role at the designated site. *See also* failover.

slave

The controller that powers up last. Also called *storage system slave*.

SMF
Single-mode fiber. A type of cable, typically 9-µm, although 8-µm and 10-µm are also supported.

Snapclone
A snapshot or virtual copy of the data, which becomes a clone or physical copy of the source over time.

snapshot
A temporary virtual disk that reflects the contents of another virtual disk at a particular point in time. A snapshot operation is performed only on an active virtual disk. The active disk and its snapshot constitute a virtual family. There are two types of snapshots: fully allocated (space inefficient) and demand-allocated (space efficient).

SONET (Synchronous Optical Network)
An ANSI standard for transmitting bits over fiber optic cable. See OC1, OC3, OC12, and OC48.

source
Used as a context-sensitive adjective, source describes the DR group, array/EVA, storage system, and so on, where an original I/O is stored before replication.

source Vdisk
A virtual disk that has the original data which is replicated to a destination virtual disk.

Storage Management Appliance
A single aggregation point for data management. The HP OpenView Storage Management Appliance is a specialized server on which SAN applications, the Element Manager, and the Continuous Access UI run.

storage pool
The aggregated blocks of available storage in the total physical disk array.

storage system
An array of HSV controllers and the array of physical disks they control. A storage system may contain virtual disks that are sources as well as virtual disks that are destinations. Sometimes used in context as follows:

- Source storage system—Used in the context of a particular data replication (DR) group, this is the storage system in which the source virtual disk resides.
- Destination storage system—Used in the context of a particular DR group, this is the storage system at which the destination virtual disk resides.

storage virtualization
The transparent abstraction of storage at the block level. It separates out logical data access from physical per disk, or per array, data access.

suspend
Command issued to a data replication (DR) group or managed set that temporarily halts replication of I/O from all source virtual disks to destination virtual disks in that DR group. Source virtual disks continue to run I/O locally, and the I/O is also copied to the DR group log. May not be issued if the DR group is failsafe enabled. *See also* resume.

synchronous mode
The mode of operation of a data replication (DR) group where the data is written to the source and destination caches, after which a completion acknowledgement is sent to the server. *See also* DR group write mode.

T1
The standard North American carrier for transmission at 1.544 Mbps.

T2
The standard North American carrier for transmission at 6.176 Mbps.

T3
The standard North American carrier for transmission at 44.736 Mbps.

throughput
In data transmission, a measure of performance for the amount of data moved successfully from one place to another in a given time.

trunking
The combining of two or more low-speed links into one virtual high-speed link. In a Fibre Channel fabric, trunking means combining two or more intersite links (ISLs) into one virtual high-speed ISL.

unspecified bit rate (UBR)
Offers no traffic-related Quality of Service (QoS) guarantees. Not acceptable for FCIP use.

UUID
Universal unique identifier. A unique 128-bit identifier associated with HSV objects.

VCS
Virtual Controller Software. The firmware that runs the storage system.

Vdisk
Virtual disk accessible from servers attached to the SAN. When it is a member of a data replication group, a virtual disk is the HSV implementation of a copy set, and it can have two states: *normal* and *copying*.

very long-distance GBIC
A GBIC used with 9-µm single-mode fiber and operating on a wavelength of 1500 nm. The 1-Gbps model has SC-style connectors and runs to a distance of 100 km (160 km with two hops). The 2-Gbps model has LC-style connectors, but is not available at the time of publication. Also known as a *VLD GBIC*.

virtual channel (VC)
The lowest-order logical address in asynchronous transfer mode of an ATM network. VC refers to a given circuit on a link.

virtual channel identifier (VCI)
The field of the cell header that stores the VC address.

Virtual Controller Software
See VCS.

virtual disk
See Vdisk.

virtual path (VP)
The highest-order part of a logical address in ATM networks. VP refers to a given group of circuits on a link.

virtual path identifier (VPI)
The field of the cell header that stores the VP address.

Vraid0
A virtualization technique that provides no data protection. The data server is divided into chunks and distributed on the disks that institute the disk group from which the virtual disk was created. Reading and writing to a Vraid0 virtual disk is very fast and makes the fullest use of the available storage, but provides no data protection (redundancy). Note that HP recommends against using Vraid0 virtual disks except for limited uses, such as nonreplicated temporary scratch disks.

Vraid1
A virtualization technique that provides the highest level of data protection. All data blocks are mirrored, or written twice, on separate physical disks. For read requests, the block can be read from either disk, which can increase performance. Mirroring requires the most storage space because twice the storage capacity must be allocated for a given amount of data.

Vraid5
A virtualization technique that uses parity striping to provide moderate data protection. Parity is a data protection mechanism for a striped virtual disk, on which the data to and from the server is broken down into chunks and distributed to the physical disks that constitute the disk group in which the virtual disk was created. If the striped virtual disk has parity, another chunk (a parity chunk) is calculated from the set of data chunks and written to the physical disks. If one of the data chunks becomes corrupted, the data can be reconstructed from the parity chunk and the remaining data chunks.

wavelength division multiplexing (WDM)
The ability to have multiple optical signals share a single optical cable.

world wide LUN name (WWLN)
The 128-bit identifier associated with a virtual disk (64 bits come from the controller WWN).

world wide name (WWN)
A 64- or 128-bit identifier used to uniquely identify the address of a component on the fabric.

WWLN
See world wide LUN name.

WWN
See world wide name.

Index

O

P

R

S